

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OH protein - protein search, using sw model

Run on: November 1, 2002, 12:14:41 : Search time 26.5 seconds  
(without alignments)  
62.872 Million cell updates/sec

Title: US-09-529-691A-1  
Perfect score: 92  
Sequence: 1 GVKGDKGNPCWPGAP 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A-Geneseq\_032802:\*

1: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	12	AA12973
2	92	100.0	15	16	AA183639
3	92	100.0	15	16	AA182922
4	92	100.0	15	17	AAW01139
5	92	100.0	15	18	AAW13022
6	92	100.0	15	19	AAW56114
7	92	100.0	15	20	AAW04481
8	92	100.0	15	21	AAW78574
9	92	100.0	24	19	AAW45044
10	92	100.0	27	18	AAW13025
11	92	100.0	39	19	AAW46046

12	92	100.0	39	19	AAW46043
13	92	100.0	43	18	AAW13024
14	92	100.0	1669	22	AAW40863
15	92	100.0	1672	22	AAW39077
16	86	93.5	406	21	AAW58169
17	69	75.0	772	13	AAW23873
18	69	75.0	772	18	AAW09643
19	69	75.0	1685	22	ABG04839
20	69	75.0	1693	22	ABG15619
21	64	69.6	430	21	AAW58180
22	61	60.3	224	22	ABW68977
23	61	66.3	429	22	AAU16066
24	61	66.3	532	19	AAW40114
25	61	66.3	1690	22	AAW23916
26	60	65.2	1078	16	AAW71704
27	60	65.2	1078	21	AAW96125
28	60	65.2	1196	13	AAW28916
29	60	65.2	1466	22	ABW50291
30	60	65.2	1466	22	AAE02533
31	60	65.2	1466	22	AAE02534
32	60	65.2	1466	22	AAE02537
33	60	65.2	1469	22	ABG15191
34	60	65.2	1940	22	ABW64070
35	59	64.1	471	16	AAW79163
36	59	64.1	471	20	AAW44171
37	59	64.1	471	21	AAW56783
38	59	64.1	471	22	AAE09483
39	59	64.1	546	22	ABG15617
40	59	64.1	547	22	ABG04840
41	59	64.1	1078	22	AAW40012
42	58	63.0	654	22	AAW63332
43	58	63.0	666	22	AAW63343
44	58	63.0	1288	20	AAW92297
45	58	63.0	1301	20	AAW92296

#### ALIGNMENTS

RESULT 1  
AA12973 standard; peptide: 15 AA.  
ID AA12973;  
AC AA12973;  
XX 24-SEP-1991 (first entry)  
DT  
XX  
XX IV-H1 based on type IV collagen alpha 1 chain triple helical region.  
DE malignant cell growth; melanoma cell motility; cellular adhesion.  
KW  
XX  
XX Synthetic.  
OS  
XX  
XX WO9108755-A.  
PN  
XX  
XX 27-JUN-1991.  
PD  
XX  
XX 06-DEC-1990; 90WO-US07162.  
PE  
XX  
XX 14-DEC-1989; 89US-0450419.  
PR  
XX  
XX (MIND) MINNESOTA UNIVERSITY.  
PA  
XX  
XX Chelberg MK, Tsilibary PE, McCarthy JB;  
PI  
XX  
XX WPI: 1991-207864/28.  
DR  
XX  
XX Polypeptide for promoting tumour cell adhesion, spreading and  
PT motility - useful in chemo:diagnosis and chemotherapy of  
PT malignant cell growth  
XX  
XX  
PS Claim 1: Page 26: 37bp: English.

CC This polypeptide (designated IV-H1) corresponds to type IV collagen  
 CC residues 1263-1277 from the major triple helical region of the alpha  
 CC 1 chain type of IV collagen. It, or a polypeptide comprising this  
 CC sequence, can be used to promote cellular attachment to substrata  
 CC or to inhibit the metastasis and invasion of tumour cells.

XX Sequence 15 AA;

Query Match 100.0%; Score 92; DB 12; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKNPGWPGAP 15  
 DB 1 GVKGDKNPGWPGAP 15

RESULT 2

AAR83639 standard; peptide; 15 AA.

XX AAR83639;

XX 11-APR-1996 (first entry)

XX Collagen IV alpha-2 chain NC1 domain peptide IV-H1 (1263-1277).

XX Collagen; type IV; alpha-2 chain; NC1 domain; fibroblast; cell;

XX proliferation; suppression; epithelial; glial; epiretinal; eye;

XX Membranes; contraction; migration; vitreoretinopathy; scarring;

XX vitreoretinal; glaucoma.

XX Synthetic.

XX WO9522979-A1.

XX 31-AUG-1995.

XX 28-FEB-1995; 95WO-US02478.

XX 27-FEB-1995; 95US-0203458.

XX 28-FEB-1994; 94US-0203458.

XX (MIND ) UNIV MINNESOTA.

XX Agarwal A, Balles M, Furcht LF, Gregerson DS, Murali S;

XX Wright MM, Skuditz AP;

XX WPI; 1995-320291/41.

XX Treating proliferative vitreo:retinopathy, retinal scarring and

XX glaucoma - using peptide derived from fibronectin, laminin or

XX collagen

XX Example 1; Page 29; 84pp; English.

XX A polypeptide which includes at least five amino acids from a

XX fragment of the fibronectin A chain (AAR83679-84), C-terminal G

XX domain of the laminin A chain (AAR83642/78) or NC1 domain of the

XX alpha-2 chain of type IV collagen (AAR83639-41) suppresses

XX fibroblast, epithelial and glial cell proliferation, contraction

XX of epiretinal membranes and cell migration within the eye. It

XX may be used in the treatment of proliferative vitreoretinopathy,

XX vitreoretinal scarring and glaucoma.

XX Sequence 15 AA;

Query Match 100.0%; Score 92; DB 16; Length 15;

Best Local Similarity 100.0%; Pred. No. 5.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKNPGWPGAP 15

|||||

DB 1 GVKGDKNPGWPGAP 15

RESULT 3

AAR82922 standard; peptide; 15 AA.

XX AAR82922;

XX 20-DEC-1995 (first entry)

XX Non-RGD, non-YISGR cancer metastasis inhibitory peptide #16.

XX Cancer metastasis; adhesive peptide; core sequence; dextran; cancer;

XX water soluble polysaccharide; metastasis; wound; immunogenicity.

XX Synthetic.

XX JP07089999-A.

XX 04-APR-1995.

XX 17-SEP-1993; 93JP-0254779.

XX 17-SEP-1993; 93JP-0254779.

XX (JAPG ) NIPPON ZEON KK.

XX WPI; 1995-167254/22.

XX Cancer metastasis inhibitive peptide derivs. - useful for inhibition

XX of cancer metastasis, healing of wounds and regulation of

XX immunogenicity.

XX Disclosure; Page 3; 6pp; Japanese.

XX The peptides AAR70472-90 and AAR82907-24 are peptide derivatives which

XX inhibit cancer metastasis. They are composed of an adhesive peptide

XX with a core sequence selected from: RGD (AAR70472-85), YIGSR

XX (AAR70486-90) or other sequence (AAR82907-24), linked to a water soluble

XX polysaccharide, preferably a water soluble dextran, at the C-terminus.

XX The peptides are useful in inhibiting cancer metastasis, healing wounds

XX and the regulation of immunogenicity.

XX Sequence 15 AA;

Query Match 100.0%; Score 92; DB 16; Length 15;

Best Local Similarity 100.0%; Pred. No. 5.2e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKNPGWPGAP 15

DB 1 GVKGDKNPGWPGAP 15

RESULT 4

AAW01139 standard; peptide; 15 AA.

XX AAW01139;

XX 18-DEC-1996 (first entry)

XX Peptide 10 for glia cell removal derived from collagen.

XX glia cell; neuron; analysis; behaviour; selective; removal.

XX laminin; fibronectin; collagen.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 15 /note="Pro-NH2"

```

XX JP08073495-A.
PN
XX
XX 19-MAR-1996.
PD
XX
XX 31-AUG-1994; 94JP-0232263.
PF
XX
XX 31-AUG-1994; 94JP-0232263.
PR
XX
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX
XX WPI: 1996-205531/21.
DR
XX
XX Selective remover for gila cells from neuronal cell cultures -
PT useful for studying behaviour of neurons in the absence of gila
PT cells.
XX
XX Example 1; Page 4; 6pp; Japanese.
PS
XX
XX AAW01135-40 are peptides, derived from laminin, fibronectin and
CC collagen, which selectively remove gila cells in the presence of neurons.
CC This enables analysis of behaviour of neurons in the absence of gila
CC cells.
XX
SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 17; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGWPGAP 15
   |||||||||||
DB 1 GVKGDKGNPGWPGAP 15

RESULT 5
AAW13022
ID AAW13022 standard; peptide; 15 AA.
XX
XX AAW13022;
AC
XX
XX 30-APR-1997 (first entry)
DT
XX
XX Collagen alpha1(IV) chain residues 1263-1277.
DE
XX
XX Collagen; cell; adhesion; migration; promotion; alpha1(IV) chain;
KW triple helical polypeptide; solid phase synthesis; study;
KW O-glycosidic bond; deuterium label; structure; biological activity.
XX
XX Homo sapiens.
OS
XX
XX US5576419-A.
PN
XX
XX 19-NOV-1996.
PD
XX
XX 30-JUN-1993; 93US-0085633.
PF
XX
XX 30-JUN-1993; 93US-0085633.
PR
XX
XX 27-SEP-1995; 95US-0534342.
PA (MINU ) UNIV MINNESOTA.
XX
XX Fields GB;
PI
XX
XX WPI: 1997-011309/01.
DR
XX
XX Solid phase synthesis of triple-helical branched polypeptide -
PT under O-glycosidic bond and deuterium label retaining conditions,
PT which may contain collagen cell adhesion sequences, useful for
PT studying structure and biological activity of collagen
XX
XX Claim 18; Column 25; 25pp; English.
XX

```

```

CC The present peptide is a collagen cell adhesion sequence (CCAS),
CC known to promote adhesion and migration of various cells,
CC comprising residues 1263-1277 of the collagen alpha1(IV) chain.
CC The triple helical polypeptide 2 (THP-2), which includes the
CC present CCAS (residues 25-39) in its triple helix inducing region,
CC was prepared by solid phase synthesis under O-glycosidic bond and
CC deuterium label retaining conditions. THP-2 can be used to study
CC the structure and biological activity of collagen.
XX
XX Sequence 15 AA;

SQ
XX
XX Query Match 100.0%; Score 92; DB 18; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGWPGAP 15
   |||||||||||
DB 1 GVKGDKGNPGWPGAP 15

RESULT 6
AAW56114
ID AAW56114 standard; peptide; 15 AA.
XX
XX AAW56114;
AC
XX
XX 09-JUL-1998 (first entry)
DT
XX
XX Peptide representing the alpha-1(IV)1263-1277 collagen sequence.
DE
XX
XX Alpha-1(IV)1263-1277 collagen sequence; peptide-amphiphile complex;
KW lipophilic portion; stable lipid film; bilayer membrane system;
KW drug-targeting; melanoma cell.
XX
XX Synthetic.
OS
XX
XX WO9807752-A1.
PN
XX
XX 26-FEB-1998.
PD
XX
XX 27-AUG-1996; 96WO-US14145.
PF
XX
XX 23-AUG-1996; 96US-0702254.
PR
XX
XX (MINU ) UNIV MINNESOTA.
PA
XX
XX Fields GB, Tirrell MV;
PI
XX
XX WPI: 1998-179051/16.
DR
XX
XX Peptide-amphiphile complexes which are able to form micelles or
PT vesicles - comprise a lipophilic portion and a peptide portion which
PT has a secondary structure
PT
XX
XX Claim 15; Page 23; 33pp; English.
PS
XX
XX The present sequence represents the alpha-1(IV)1263-1277 collagen
CC sequence. This sequence is used to construct a peptide-amphiphile
CC complex to exemplify the invention. The peptide-amphiphile complex of
CC the specification comprises a lipophilic portion and a peptide portion
CC having a secondary structure. The peptide portion preferably has
CC biological activity, e.g. cell recognition activity or enzymatic
CC activity. The complex may be used for forming stable lipid films on
CC substrates, or may be used in bilayer membrane systems. It may, in
CC micellar or vesicular form, be used in preparation of drug-targeting
CC systems against melanoma cells.
XX
XX Sequence 15 AA;

SQ
XX
XX Query Match 100.0%; Score 92; DB 19; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GVGKDGKGNPMPGAP 15  
 |||||  
 DB 1 GVGKDGKGNPMPGAP 15

RESULT 7  
 ID AAY04481  
 AAY04481 standard; peptide; 15 AA.

XX AAY04481;

XX 05-JUL-1999 (first entry)

XX Human type IV collagen alpha 1 chain D form fragment D-IVH1.

XX Human; type IV collagen alpha 1 chain; D-IVH1; inhibition; tumour;

KW cell adhesion; major triple helical domain; collagenous region.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1..15 /note="D-form residues"

XX WO9920300-A1.

XX 29-APR-1999.

XX 22-OCT-1998; 98WO-US22405.

XX 22-OCT-1997; 97US-0062716.

PR 22-OCT-1997; 97US-0062617.

XX (MINU ) UNIV MINNESOTA.

XX Fields GB, McCarthy JB;

XX WPI; 1999-302644/25.

XX New polypeptide useful for inhibition of tumour cell adhesion to type

PT IV collagen

XX Claim 4; Page 18; 31pp; English.

XX The present sequence represents an all D-form fragment of the continuous

CC collagenous region of the major triple helical domain of the alpha 1

CC chain of human type IV collagen, designated D-IVH1. The peptide D-IVH1

CC can be used to inhibit tumour cell binding to type IV collagen, inhibit

CC tumour cell invasion of basement membrane and inhibit tumour cell

CC metastasis in vivo. The peptide can be used to treat melanoma cell

CC metastasis, and spontaneous Lewis lung tumour metastasis. The peptide is

CC also useful for targeting cytotoxic agents and drugs to tumour cells.

XX Sequence 15 AA;

QY 1 GVGKDGKGNPMPGAP 15  
 |||||  
 DB 1 GVGKDGKGNPMPGAP 15

RESULT 8  
 ID AAY78574  
 AAY78574 standard; peptide; 15 AA.

XX AAY78574;

XX 05-MAY-2000 (first entry)

DE Type IV collagen derived peptide IV-H1.  
 XX Type IV collagen; intraocular fibroblast proliferation inhibitor;  
 KW proliferative retinopathy; fibroblast migration; scarring; glioma.

XX Unidentified.

XX US6013628-A.

XX 11-JAN-2000.

XX 27-FEB-1995; 95US-0394748.

XX 28-FEB-1994; 94US-0203458.

XX (MINU ) UNIV MINNESOTA.

PI Gregerson DS, Agarwal A, Wright MM, Murali S, Skuditz APN;

PI Furcht LT, Balle M;

XX WPI; 2000-159882/14.

XX Treatment of glaucoma by ocular administration of a type IV collagen

PT peptide that inhibits fibroblast proliferation

XX Example 1; Column 15; 44pp; English.

XX This sequence represents a type IV collagen peptide. Ocular

CC administration of the peptide inhibits fibroblast proliferation and can

CC be used in the treatment of glaucoma. The peptide can be used to treat

CC proliferative vitreoretinopathy by ocular administration of the peptides

CC of the invention, which suppress fibroblast-mediated collagen gel

CC contraction. The peptides can be used for treating glaucoma in mammals,

CC especially humans, by inhibiting intraocular fibroblast proliferation,

XX fibroblast migration and scarring.

XX Sequence 15 AA;

Query Match 100.0%; Score 92; DB 21; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVGKDGKGNPMPGAP 15  
 |||||  
 DB 1 GVGKDGKGNPMPGAP 15

RESULT 9  
 ID AAW46044  
 AAW46044 standard; peptide; 24 AA.

XX AAW46044;

XX 23-DEC-1998 (first entry)

XX Collagen sequence synthesised as triple-helical peptide.

XX Triple helix; collagen structure; homo-trimer; hetero-trimer.

XX Synthetic.

XX Key Location/Qualifiers

XX Modified-site 3 /label= 4Hyp

XX Modified-site 6 /label= 4Hyp

XX Modified-site 9 /label= 4Hyp

XX US5726243-A.

XX 10-MAR-1998.



```

PF 03-JUL-1996; 96US-0675140.
XX
PR 30-JUN-1993; 93US-0085633.
PR 27-SEP-1995; 95US-0534342.
PR 03-JUL-1996; 96US-0675140.
XX
PA (MINU ) UNIV MINNESOTA.
XX
PI Fields GB;
XX
DR WPI; 1998-192815/17.
XX
PT Supported polypeptide(s) useful as intermediates - for
PT triple-helical polypeptide(s) having collagen structures
XX
PS Disclosure; Column 3; 23pp; English.
XX
XX The invention relates to supported polypeptides of formula:
CC A-B-(J)-[U-L-R, where A = an N alpha-amino protecting group
CC removable under non-acidic conditions; B = an amino acid having a single
CC side-chain amino group protected with a group removable under non-strong
CC acidic conditions; J = an amino acid that is capable of acting as a
CC chromophore and is protected with a group capable of withstanding the
CC conditions under which the protecting groups of A and B are removed; U =
CC an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
CC labilised using a non-strong acid mechanism, and R = a support material.
CC The supported polypeptides are intermediates for triple-helical
CC polypeptides. The triple-helical polypeptides are useful for studying the
CC structure and biological activity of specific collagen sequences in homo-
CC and hetero-trimeric forms. The present sequence represents a collagen
CC sequence used in the synthesis of triple-helical peptides.
XX
SQ Sequence 24 AA:
OY 1 GVGKDGKNGMPGAP 15
DB 10 GVGKDGKNGMPGAP 24

RESULT 10
AAW13025
ID AAW13025 standard; peptide: 27 AA.
XX
XX AAW13025;
AC
XX
DT 30-APR-1997 (first entry)
XX
DE Collagen alpha1(IV) chain containing triple helical polypeptide 4.
XX
XX Triple helical polypeptide; collagen; cell; adhesion; migration;
XX promotion; alpha(IV) chain; solid phase synthesis; study;
XX O-glycosidic bond; deuterium label; structure; biological activity;
XX homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 3 /label- Hyp
XX Modified-site 6 /label- Hyp
XX Modified-site 9 /label- Hyp
XX Modified-site 10..24 /label- Hyp
XX Peptide /note- "collagen alpha1(IV) chain residues
XX 1263-1277"
XX Modified-site 25 /note- "linked via peptide bond to Pro24 of peptide
XX comprising residues 1-24 of present

```

```

FT Modified-site 26 sequence"
FT /note- "linked via peptide bond to Pro24 of peptide
FT comprising residues 1-24 of present
FT sequence"
XX
XX US5576419-A.
XX
XX 19-NOV-1996.
XX
XX 30-JUN-1993; 93US-0085633.
XX
XX 30-JUN-1993; 93US-0085633.
XX
XX 27-SEP-1995; 95US-0534342.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Fields GB;
XX
XX WPI; 1997-011309/01.
XX
XX Solid phase synthesis of triple-helical branched polypeptide -
XX under O-glycosidic bond and deuterium label retaining conditions,
XX which may contain collagen cell adhesion sequences, useful for
XX studying structure and biological activity of collagen
XX
XX Claim 52; Column 27; 25pp; English.
XX
XX The present sequence is that of triple helical polypeptide 4
XX (THP-4), which includes a collagen cell adhesion sequence, known to
XX promote adhesion and migration of various cells (residues 1263-1277
XX of the collagen alpha1(IV) chain), in its triple helix inducing
XX region. THP-4 was prepared by solid phase synthesis under
XX O-glycosidic bond and deuterium label retaining conditions, without
XX strong acidolysis steps which affect O-glycosylation and isotopic
XX labelling. THP can be used to study the structure and biological
XX activity of particular collagens as homotrimeric or heterotrimeric
XX forms, with the advantage that different adhesion sequences can be
XX studied with a minimum of Gly-Pro-Hyp triplets, avoiding
XX overlapping NMR resonances.
XX
XX Sequence 27 AA:
SQ
OY 1 GVGKDGKNGMPGAP 15
DB 10 GVGKDGKNGMPGAP 24

RESULT 11
AAW46046
ID AAW46046 standard; peptide: 39 AA.
XX
XX AAW46046;
AC
XX
DT 23-DEC-1998 (first entry)
XX
XX Peptide intermediate used in synthesis of triple-helical peptide.
XX Triple helix; collagen structure; homo-trimer; hetero-trimer.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Modified-site 3 /label- 4Hyp
XX Modified-site 6 /label- 4Hyp
XX Modified-site 9 /label- 4Hyp
XX Modified-site 9 /label- 4Hyp

```

```

FT Modified-site 12 /label= 4Hyp
FT Modified-site 15 /label= 4Hyp
FT Modified-site 18 /label= 4Hyp
FT Modified-site 21 /label= 4Hyp
FT Modified-site 24 /label= 4Hyp
FT Modified-site 27 /label= 4Hyp
FT Modified-site 29 /note= "Lys(Boc)"
FT Modified-site 30 /note= "Asp(OtBu)"
FT Modified-site 32 /note= "Lys(Boc)"
FT Modified-site 32 /note= "Asn(Trt)"
XX
XX US5726243-A.
XX
XX 10-MAR-1998.
XX
XX 03-JUL-1996; 96US-0675140.
XX
XX 30-JUN-1993; 93US-0085633.
XX 27-SEP-1995; 95US-0534342.
XX 03-JUL-1996; 96US-0675140.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Fields GB;
XX
XX WPI; 1998-192815/17.
XX
XX Supported polypeptide(s) useful as intermediates - for
XX triple-helical polypeptide(s) having collagen structures
XX
XX Disclosure; Figure 3; 23pp; English.
XX
XX The invention relates to supported polypeptides of formula:
XX A-B-B-(U)j-(U)u-L-R; where A = an N alpha-amino protecting group
XX removable under non-acidic conditions; B = an amino acid having a single
XX side-chain amino group protected with a group removable under non-strong
XX acidic conditions; j = an amino acid that is capable of acting as a
XX chromophore and is protected with a group capable of withstanding the
XX conditions under which the protecting groups of A and B are removed; u =
XX an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
XX labilised using a non-strong acid mechanism, and R = a support material.
XX The supported polypeptides are intermediates for triple-helical
XX polypeptides. The triple-helical polypeptides are useful for studying the
XX structure and biological activity of specific collagen sequences in homo-
XX and hetero-trimeric forms. The present sequence represents a peptide
XX intermediate used in the synthesis of a triple-helical peptide.
XX
XX Sequence 39 AA;
XX
XX Query Match 100.0%; Score 92; DB 19; Length 39;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GVGDKGNPGWPGAP 15
DB 25 GVGDKGNPGWPGAP 39
RESULT 12
AAW46043
XX AAW46043 standard; peptide; 39 AA.
XX
XX AAW46043;
XX
XX 23-DEC-1998 (first entry)

```

```

XX XX Collagen sequence synthesised as triple-helical peptide.
XX DE Triple helix; collagen structure; homo-trimer; hetero-trimer.
XX XX Synthetic.
XX KW Key
XX OS Location/Qualifiers
XX
XX FH Modified-site 3 /label= 4Hyp
XX FT Modified-site 6 /label= 4Hyp
XX FT Modified-site 9 /label= 4Hyp
XX FT Modified-site 12 /label= 4Hyp
XX FT Modified-site 15 /label= 4Hyp
XX FT Modified-site 18 /label= 4Hyp
XX FT Modified-site 21 /label= 4Hyp
XX FT Modified-site 24 /label= 4Hyp
XX FT Modified-site 24 /label= 4Hyp
XX
XX US5726243-A.
XX
XX 10-MAR-1998.
XX
XX 03-JUL-1996; 96US-0675140.
XX
XX 30-JUN-1993; 93US-0085633.
XX 27-SEP-1995; 95US-0534342.
XX 03-JUL-1996; 96US-0675140.
XX
XX (MINU ) UNIV MINNESOTA.
XX
XX Fields GB;
XX
XX WPI; 1998-192815/17.
XX
XX Supported polypeptide(s) useful as intermediates - for
XX triple-helical polypeptide(s) having collagen structures
XX
XX Disclosure; Column 3; 23pp; English.
XX
XX The invention relates to supported polypeptides of formula:
XX A-B-B-(U)j-(U)u-L-R; where A = an N alpha-amino protecting group
XX removable under non-acidic conditions; B = an amino acid having a single
XX side-chain amino group protected with a group removable under non-strong
XX acidic conditions; j = an amino acid that is capable of acting as a
XX chromophore and is protected with a group capable of withstanding the
XX conditions under which the protecting groups of A and B are removed; u =
XX an amino acid; u = 0 or 1; j = 1 or more; L = a linker capable of being
XX labilised using a non-strong acid mechanism, and R = a support material.
XX The supported polypeptides are intermediates for triple-helical
XX polypeptides. The triple-helical polypeptides are useful for studying the
XX structure and biological activity of specific collagen sequences in homo-
XX and hetero-trimeric forms. The present sequence represents a collagen
XX sequence used in the synthesis of triple-helical peptides.
XX
XX Sequence 39 AA;
XX
XX Query Match 100.0%; Score 92; DB 19; Length 39;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-06;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GVGDKGNPGWPGAP 15
DB 25 GVGDKGNPGWPGAP 39
RESULT 13

```

AAW13024  
ID AAW13024 standard; peptide: 43 AA.  
XX  
AC AAW13024;  
XX  
DT 30-APR-1997 (first entry)  
XX  
DE Collagen alpha1(IV) chain containing triple helical polypeptide 2.  
XX  
KM Triple helical polypeptide; collagen; cell; adhesion; migration;  
KM promotion; alpha1(IV) chain; solid phase synthesis; study;  
KW O-glycosidic bond; deuterium label; structure; biological activity;  
KW homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT Modified-site 3  
FT /label- Hyp 6  
FT Modified-site 6  
FT /label- Hyp 9  
FT Modified-site 9  
FT /label- Hyp 12  
FT Modified-site 12  
FT /label- Hyp 15  
FT Modified-site 15  
FT /label- Hyp 18  
FT Modified-site 18  
FT /label- Hyp 21  
FT Modified-site 21  
FT /label- Hyp 24  
FT Modified-site 24  
FT /label- Hyp 25..39  
FT Peptide /note- "collagen alpha1(IV) chain residues 1263-1277"  
FT Modified-site 40  
FT /note- "linked via peptide bond to Pro39 of peptide comprising residues 1-39 of present sequence"  
FT Modified-site 41  
FT /note- "linked via peptide bond to Pro39 of peptide comprising residues 1-39 of present sequence"  
XX  
PN US5576419-A.  
XX  
PD 19-NOV-1996.  
XX  
PF 30-JUN-1993; 93US-0085633.  
XX  
PR 30-JUN-1993; 93US-0085633.  
PR 27-SEP-1995; 95US-0534342.  
XX  
PA (MINU ) UNIV MINNESOTA.  
XX  
PI Fields GB;  
XX  
DR WPI; 1997-011309/01.  
XX  
PT Solid phase synthesis of triple-helical branched polypeptide -  
PT under O-glycosidic bond and deuterium label retaining conditions,  
PT which may contain collagen cell adhesion sequences, useful for  
PT studying structure and biological activity of collagen  
XX  
PS Claim 48; Column 27; 25pp; English.  
XX  
CC The present sequence is that of triple helical polypeptide 2  
CC (THP-2), which includes a collagen cell adhesion sequence, known to  
CC promote adhesion and migration of various cells (residues 1263-1277  
CC of the collagen alpha1(IV) chain), in its triple helix inducing  
CC region. THP-2 was prepared by solid phase synthesis under  
CC O-glycosidic bond and deuterium label retaining conditions, without

CC strong acidolysis steps which affect O-glycosylation and isotopic  
CC labelling. THP can be used to study the structure and biological  
CC activity of particular collagens as homotrimeric or heterotrimeric  
CC forms, with the advantage that different adhesion sequences can be  
CC studied with a minimum of Gly-Pro-Hyp triplets, avoiding  
CC overlapping NMR resonances.  
XX  
SQ Sequence 43 AA:  
XX  
QY 1 GVGDKGNPWPAP 15  
DB 25 GVGDKGNPWPAP 39  
XX  
RESULT 14  
AAW13024  
ID AAW13024 standard; protein: 1669 AA.  
XX  
AC AAW13024;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 5794.  
XX  
XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
OS Homo sapiens.  
XX  
PN W0200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0596042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HYSE-) HYSEO INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao Q, Zhou P, Goodrich R, Dymann RT;  
XX  
DR WPI; 2001-442253/47.  
XX  
DR N-PSDB; AAI60019.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Example 2; SEQ ID NO 5794; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAW13024-AAW13024) with noctropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as



CC chromosome identification, as chromosome markers, and for numerous other  
CC diagnostic or research purposes. The proteins may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders. The proteins may also be used in the treatment of wounds and  
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
CC peptide AAB58549 are used in the course of the invention for the  
CC identification and characterization of the polynucleotide and protein  
CC sequences.

XX Sequence 406 AA;

Query Match 93.5%; Score 86; DB 21; Length 406;  
Best Local Similarity 100.0%; Pred. No. 0.00013;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 VKGDKNPGWPGAP 15  
DB 1 VKGDKNPGWPGAP 14

## RESULT 17

AA023873  
ID AAR23873 standard; Protein: 772 AA.

XX AAR23873;

XX 25-NOV-1992 (first entry)

DE Human alpha 5 (IV) of type IV collagen.

XX Mutations; Alport's syndrome; basement membranes; diabetes mellitus.

XX Homo sapiens.

OS Location/Qualifiers

FT Key M1sc-difference 43..47 /note="Interruption in Gly-X-Y sequence"

FT M1sc-difference 159..160 /note="Interruption in Gly-X-Y sequence"

FT M1sc-difference 275..277 /note="Interruption in Gly-X-Y sequence"

FT M1sc-difference 334..336 /note="Interruption in Gly-X-Y sequence"

FT M1sc-difference 456..458 /note="Interruption in Gly-X-Y sequence"

XX US5114840-A.

XX 19-MAY-1992.

XX 07-JUL-1989; 89US-0377238.

XX 07-JUL-1989; 89US-0377238.

XX (TRYG/) TRYGVAASON K.

XX Hostikka SL, Tryggvason K;

XX WPI: 1992-192174/23.

XX N-PSDB; AAO24551.

XX Isolation of DNA encoding alpha-5(IV)polypeptide of type IV

XX collagen - to detect mutations in genes for alpha-5(IV) chain

XX which produce genetic or acquired basement membrane disorders

XX e.g. Alport's syndrome

XX Disclosure: Fig 2; 14pp; English.

XX The sequence is that of the alpha 5(IV) polypeptide chain of human

XX type IV collagen, the major component of basement membranes. The

XX collagen contains the Gly-X-Y repeat coding sequence typical for

XX collagenous proteins at one end and a typical NC-domain coding

CC sequence at the other end. The sequence can be used to detect  
CC mutations in individual genes specific for this chain which can,  
CC directly or indirectly, produce several human diseases. It can  
CC also be used to determine genetic e.g. Alport's syndrome, or  
CC acquired e.g. diabetes mellitus, disorders of the basement membrane,  
CC and as probes or antibodies against these nucleotide sequences.  
CC Gene fragments generated through amplifications from human genomic  
CC or cloned DNA can also be used for detection and analysis of genes.

XX Sequence 772 AA;

Query Match 75.0%; Score 69; DB 13; Length 772;  
Best Local Similarity 73.3%; Pred. No. 0.074;

Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVGDKNPGWPGAP 15  
DB 366 GVGDKNPGWPGAP 380

## RESULT 18

AA009643  
ID AAW09643 standard; Protein: 772 AA.

XX AAW09643;

XX 16-JUN-1997 (first entry)

DE Human type IV collagen alpha-5.

XX Collagen alpha5(IV); basement membrane; Alport's syndrome;

XX nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene;

XX X chromosome.

XX Homo sapiens.

OS Location/Qualifiers

FT Key Domain 1..543 /label="Collagenous domain

FT M1sc-difference 43..47, 159..160, 275..276, 334..335, 456..459"

FT Domain 544..772 /label="Non-collagenous domain

FT Peptide 742..751 /label="Immunogenic peptide

FT M1sc-difference 742..751 /note="peptide used to raise diagnostic

XX US5593900-A.

XX 14-JAN-1997.

XX 07-JUL-1989; 89US-0377238.

XX 20-DEC-1990; 90US-0630563.

XX 07-JUL-1989; 89US-0377238.

XX 11-OCT-1994; 94US-0321084.

XX (HOST/) HOSTIKKA S L.

XX (HOYT/) HOYTIVA M.

XX (TRYG/) TRYGVAASON K.

XX Hostikka SL, Hoytva M, Tryggvason K;

XX WPI: 1997-099481/09.

XX N-PSDB; AAT47812.

XX New antibodies specific for human type IV collagen alpha5 chain -

XX used to detect absence of this chain in patients with renal failure

XX Disclosure: Fig 2A-2B; 12pp; English.

XX The amino acid sequence of a portion (AAW09643) of the previously  
 CC unknown human type IV collagen chain, alpha5(IV), was deduced from  
 CC cDNA clones (see also AAT47812) obtd. using probes based on  
 CC conserved sequences of human alpha1(IV) and alpha2(IV) collagen  
 CC chains and of the Drosophila alpha(IV) chain. It includes a  
 CC complete non-collagenous domain that shows 83% identity with that  
 CC of alpha1(IV) and 63% with that of the alpha2(IV) chain. Mutations  
 CC in the alpha5(IV) gene (COL4A5) are associated with Alport's  
 CC syndrome. Antibodies raised against a peptide (see also AAW09644)  
 CC specific to alpha5(IV) can be used in the diagnosis of basement  
 CC membrane disorders such as Alport's syndrome.  
 XX

SQ Sequence 772 AA;  
 Query Match 75.0%; Score 69; DB 18; Length 772;  
 Best Local Similarity 73.3%; Pred. No. 0.074;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKNPGWPGAP 15  
 :||:||||| ||  
 Db 366 GIKGEKGNPGQGLP 380

RESULT 19  
 ABG04839  
 ID ABG04839 standard; Protein: 1685 AA.  
 XX  
 AC ABG04839;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #4830.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS69026.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 PS Claim 20; SEQ ID No 35198; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

SQ Sequence 1685 AA;  
 Query Match 75.0%; Score 69; DB 22; Length 1685;  
 Best Local Similarity 73.3%; Pred. No. 0.17;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKNPGWPGAP 15  
 :||:||||| ||  
 Db 1279 GIKGEKGNPGQGLP 1293

RESULT 20  
 ABG15619  
 ID ABG15619 standard; Protein: 1693 AA.  
 XX  
 AC ABG15619;  
 XX  
 DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #15610.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS79806.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 PS Claim 20; SEQ ID No 45978; 103pp; English.  
 XX  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1693 AA;  
Query Match Best Local Similarity 75.0%; Score 69; DB 22; Length 1693;  
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 GVKGDKNGMPGAP 15  
DB 1287 GIKGKGNPGLP 1301  
RESULT 21  
AAB58180 ID AAB58180 standard; Protein; 430 AA.  
XX  
AC AAB58180;  
XX  
DT 14-MAR-2001 (first entry)  
XX  
DE Lung cancer associated polypeptide sequence SEQ ID 518.  
XX  
XX Human; lung cancer associated protein; neuroprotective; cytostatic;  
KM cardiocative; immunomodulatory; muscular active; vulnerary;  
KM gastrointestinal; nephrotoxic; antineoplastic; gynecologically;  
KM antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
KM proliferative disorder; wound healing; infectious disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200055180-A2.  
XX  
PD 21-SEP-2000.  
XX  
PE 08-MAR-2000; 2000WO-US05918.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCT INC.  
PA (ROSE/) ROSEN C A.  
XX  
PI Ruben SM;  
XX  
PI WPI: 2000-587514/55.  
DR N-PSDB; AAF18056.  
XX  
XX Lung cancer associated gene sequences, referred to as lung cancer  
PT antigens, useful for treatment, prevention, and diagnosis of disorders  
PT such as lung cancer -  
XX  
XX Claim 11; Page 1008-1010; 1425pp; English.  
XX  
PS Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer  
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
CC associated proteins and polynucleotide sequences, their agonists, and  
CC antagonists may have neuroprotective; cytostatic; cardioactive;  
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal  
CC general; nephrotoxic; antineoplastic; gynecologically; or antibacterial  
CC activity. The invention also includes antibodies specific for the  
CC protein or polynucleotide sequences. The lung cancer associated  
CC polynucleotide sequences may be used for detection of lung cancer,  
CC chromosome identification, as chromosome markers, and for numerous other  
CC diagnostic or research purposes. The proteins may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders. The proteins may also be used in the treatment of wounds and  
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and  
CC peptide AAB58549 are used in the course of the invention for the  
CC identification and characterisation of the polynucleotide and protein  
CC sequences.  
XX  
SQ Sequence 430 AA;  
Query Match Best Local Similarity 69.6%; Score 64; DB 21; Length 430;  
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 GVKGDKNGMPGAP 15  
DB 86 GPKGKGNPGLP 100  
RESULT 22  
AAB68977 ID AAB68977 standard; Protein; 224 AA.  
XX  
AC AAB68977;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 33723.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KM pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, LI PWD, Myers EW;  
XX  
DR WPI: 2001-656860/75.  
DR N-PSDB; ABL13080.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 33723; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (AAB57737-ABR2072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 224 AA;  
Query Match Best Local Similarity 66.3%; Score 61; DB 22; Length 224;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
OY 1 GVKGDKNGMPGAP 15

Db 118 GPRGDKGHPMGIP 132

RESULT 23

ID AAU16066 standard; Protein; 429 AA.

AAU16066;

07-NOV-2001 (first entry)

Human novel secreted protein, Seq ID 1019.

Human; immunosuppressive; antiarthritic; antineumatic;  
cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
neuroprotective; antibacterial; virucide; fungicide; opthalmological;  
vulnerary; secreted protein; rheumatoid arthritis;  
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
corneal infection; wound healing; epithelial cell proliferation;  
skin aging; food additive; preservative; antiproliferative.

OS Homo sapiens.

XX MO200155322-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01341.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 26-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239936.

PR 20-OCT-2000; 2000US-0240937.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.



```

PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254957.
PR 05-JAN-2001; 2001US-0255678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-488783/53.
DR N-PSDB: AAS26053.
XX
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
XX
XX Claim 11: SEQ ID NO 1019; 980bp; English.
XX
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. Antibodies to the proteins can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease. Infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection, and many other
CC disorders listed in the specification. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention.

Query Match 66.3%; Score 61; DB 22; Length 429;
Best Local Similarity 66.7%; Pred. NO. 0.58;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNRGPAGP 15
DB 23 GAKGEGNAGFPGLP 37

RESULT 24
AAM40114
ID AAM40114 standard; Protein: 532 AA.

```

XX	AA040114;	
AC		
XX	03-JUN-1998 (first entry)	
DT		
XX		
DE	Human alpha-1(IV) collagen protein.	
XX		
KW	COL4A6 gene; type IV collagen protein; alpha-6(IV); alpha-1(IV);	
KW	Alports syndrome; leiomyomatosis; diagnosis; gene therapy; antibody.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	Region	163..171
FT		/label= I
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	224..225
FT		/label= II
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	244..257
FT		/label= III
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	273..274
FT		/label= IV
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	335..337
FT		/label= V
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	377..378
FT		/label= VI
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	403..404
FT		/label= VII
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	429..236
FT		/label= VIII
FT		/note= "This region interrupts a collagenous repeat"
FT	Region	467..473
FT		/label= IX
FT		/note= "This region interrupts a collagenous repeat"
PN	US5731192-A.	
XX		
PD	24-MAR-1998.	
XX		
PF	23-JUN-1995;	95US-0494168.
XX		
PR	27-AUG-1993;	93US-0112465.
XX		
PA	(UYA ) UNIV YALE.	
PI	Reeders ST, Zhou J;	
DR	WPI: 1998-216495/19.	
XX		
PT		
XX		
PS	Example 1; Fig 3; 43pp; English.	
XX		
CC	This sequence is the human alpha-1(IV) collagen protein which is used to analyse a novel human alpha-6(IV) collagen protein encoded by the COL4A6 gene. The alpha-6(IV) protein can be used for diagnosis or gene therapy of diseases associated with collagen type IV pathology,	
CC	especially Alport's syndrome and associated diffuse leiomyomatosis. The polypeptide may also be used for generating monoclonal or polyclonal antibodies having specificity for the alpha-6(IV) polypeptide especially an antibody that is not crossreactive with other collagen proteins including alpha-1(IV), alpha-2(IV) and alpha-5(IV) collagens.	
CC		
XX		
50	Sequence	532 AA;

Query Match

66.3%; Score 61; DB 19; Length 532;

Best Local Similarity 66.7%; Pred. No. 0.73;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVGKDKNGPWGAP 15  
Db 293 GDKGKSGSPRFP 307

RESULT 25  
ID AAM23916 standard; Protein; 1690 AA.

AC AAM23916;

DT 12-OCT-2001 (first entry)

DE Human EST encoded protein SEQ ID NO: 1441.

XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;

KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;

KM diagnostics; forensic test; gene mapping; genetic disorder;

XX biodiversity; gene therapy; nutrition.

OS Homo sapiens.

PN WO200154477-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US02687.

PR 25-JAN-2000; 2000US-0491404.

PR 17-JUL-2000; 2000US-0617746.

PR 03-AUG-2000; 2000US-0631451.

PR 15-SEP-2000; 2000US-0663870.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;

PI Cao Y, Drmanac RA, Zhang J, Werhman T;

DR WPI: 2001-476164/51.

DR N-PSDB; AAH98575.

XX isolated polypeptide for treatment of diseases, diagnostics, raising

PT antibodies and research use -

PS Claim 20; Page 1002-1006; 1275pp; English.

XX The present invention provides the protein and coding sequences of novel

CC proteins from a variety of organisms, including human, dog, cat, horse,

CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)

CC from the organism of interest. They can be used in diagnostics,

CC forensics, gene mapping, identification of mutations, to assess

CC biodiversity and for nutritional purposes. The present sequence is a

XX protein of the invention.

XX Sequence 1690 AA;

Query Match 66.3%; Score 61; DB 22; Length 1690;

Best Local Similarity 60.0%; Pred. No. 2.5;

Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVGKDKNGPWGAP 15

Db 184 GVGKDKNGPWGAP 198

RESULT 26

AAAR71704  
ID AAR71704 standard; protein; 1078 AA.

AC AAR71704;  
XX 17-OCT-1995 (first entry)

DE Collagen alpha 1 (III) chain precursor.

XX Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;

KW disorder; osteoporosis; metastatic progression; Paget's disease;

KM hyperthyroidism; bone; resorption; rheumatoid arthritis;

XX osteoarthritis; vasculitis syndrome.

OS Homo sapiens.

PN WO9508115-A.

PD 23-MAR-1995.

PF 19-SEP-1994; 94WO-DK00348.

PR 17-SEP-1993; 93DK-0001040.

XX (OSTE-) OSTEOMETER AS.

PI Bonde M, Qvist P;

DR WPI: 1995-131456/17.

PT Assaying collagen fragments in body fluid by immunoassay - using

PT antibodies raised against synthetic peptide(s) contg. potential

PT crosslinking sites, to diagnose and monitor disorders of collagen

PT metabolism, e.g. osteoporosis.

XX Disclosure (Appendix A); Page 55; 87pp; English.

CC Determination of collagen fragments in body fluids can be achieved

CC by immunoassay using antibodies directed against synthetic peptides

CC derived from collagen which contain sites of potential crosslinking.

CC The method is used to diagnose and monitor treatment of disorders of

CC collagen metabolism (degradation of type I collagen may indicate

CC osteoporosis, metastatic progression, Paget's disease,

CC hyperthyroidism or other conditions involving excessive bone

CC resorption; degradation of type II collagen may indicate rheumatoid

CC arthritis or osteoarthritis; and of type III collagen, vacuolitis

CC syndrome). The method can also be used to assess the toxicity of a

CC compound and to test drugs for their effect on collagen metabolism.

XX Sequence 1078 AA;

Query Match 65.2%; Score 60; DB 16; Length 1078;

Best Local Similarity 71.4%; Pred. No. 2.2;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGKDKNGPWGAP 14

Db 711 GVGKDKNGPWGAP 724

RESULT 27

AAAY96125  
ID AAY96125 standard; Peptide; 1078 AA.

AC AAY96125;

DT 19-DEC-2000 (first entry)

XX Collagen type III alpha-1.

XX Collagen type III; vasculitis syndrome; assay; diagnosis.

XX Homo sapiens.

XX US6110689-A.

PD 29-AUG-2000.  
 XX  
 PE 04-NOV-1997; 97US-0963825.  
 XX  
 PR 21-JAN-1994; 94US-0187319.  
 XX  
 PA (OSTE-) OSTEOMETER AS.  
 XX  
 PI Bonde M, Qvist P;  
 XX WPI: 2000-586349/55.  
 DR  
 XX Assaying type I collagen fragments for diagnosing osteoporosis in  
 PT postmenopausal women. Involves contacting body fluid with synthetic  
 PT collagen peptide and antibody and quantifying by competitive binding  
 PT assay -  
 XX  
 PS Disclosure; Column 46-51; 41pp; English.  
 XX  
 CC The present sequence is that of human type III collagen alpha-1.  
 CC The invention is based on the discovery of the presence of  
 CC particular collagen fragments in body fluids of patients compared  
 CC with those of healthy subjects. These fragments are generated  
 CC upon collagen degradation and are partly characterised by the  
 CC presence of potential sites for crosslinking. A method for  
 CC assaying collagen fragments in a body fluid sample is based on the  
 CC competitive binding to immunological binding partners of collagen  
 CC fragments in the sample and of synthetic peptides derived from  
 CC collagen and containing crosslinkable sites (see AAY9618-21). When  
 CC considering the degradation of type III collagen, the assay can be  
 CC used as a means of identifying the presence of vasculitis syndrome.  
 CC  
 SO Sequence 1078 AA;  
 OY 1 GYKDGKNGMPGA 14  
 DB 711 GVGGRGSGPGCA 724  
 |||||::|:| | | |  
 Query Match 65.2%; Score 60; DB 21; Length 1078;  
 Best Local Similarity 71.4%; Pred. No. 2.2;  
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

RESULT 28  
 AAR28916  
 ID AAR28916 standard; Protein: 1196 AA.  
 XX  
 AC AAR28916;  
 XX  
 DT 24-MAR-1993 (first entry)  
 XX  
 DE Type III procollagen (prior art).  
 XX  
 KW Mutation; pro-alpha1(III); primer; PCR.  
 KW  
 OS Homo sapiens.  
 XX  
 FN WO9219754-A.  
 PD 12-NOV-1992.  
 XX  
 PE 08-MAY-1992; 92WO-US03866.  
 XX  
 PR 08-MAY-1991; 91US-0696607.  
 XX  
 PA (UYJE-) UNIV JEFFERSON THOMAS.  
 XX  
 PI Kulvanleml SH, Prockop DJ, Tromp GC;  
 XX WPI: 1992-398878/48.  
 DR N-PSDB; AAQ30849.  
 XX  
 PT Kit for detecting genetic pre-disposition for vascular aneurysms

PT - contains primer to amplify portions of Type III procollagen DNA  
 PT and detects mutation in standard procollagen DNA  
 XX  
 PS Disclosure; Fig 1A-F; 44pp; English.  
 XX  
 CC Example 1 describes the determination of the presence of a mutation  
 CC in the pro-alpha1(III) gene. Primers used in PCR are given in  
 CC AAQ30834-48.  
 CC  
 SO Sequence 1196 AA;  
 OY 1 GYKDGKNGMPGA 14  
 DB 858 GVGGRGSGPGCA 871  
 |||||::|:| | | |  
 Query Match 65.2%; Score 60; DB 13; Length 1196;  
 Best Local Similarity 71.4%; Pred. No. 2.4;  
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

RESULT 29  
 ABB50291  
 ID ABB50291 standard; Protein: 1466 AA.  
 XX  
 AC ABB50291;  
 XX  
 DT 08-FEB-2002 (first entry)  
 XX  
 DE Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.  
 XX  
 KW Ovarian tumour marker gene; human; overexpression; upregulation;  
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
 KW identification; serous cystadenoma; borderline serous tumour;  
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KW immune response pathway; cell proliferation regulation; protein folding;  
 KW membrane localised; secreted; therapeutic target; cytostatic;  
 KW gene therapy; vaccine.  
 KW  
 OS Homo sapiens.  
 XX  
 FN WO200175177-A2.  
 PD 11-OCT-2001.  
 XX  
 PE 03-APR-2001; 2001WO-US10947.  
 XX  
 PR 03-APR-2000; 2000US-194336P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA  
 PI Morfin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 XX WPI: 2001-626450/72.  
 DR N-PSDB; ABA83117.  
 XX  
 PT Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian  
 PT cancer treatment, by measuring expression level of ovarian tumor marker  
 PT gene -  
 XX  
 PS Claim 23; Page 114-117; 140pp; English.  
 XX  
 CC The invention relates to methods for diagnosing and prognosing ovarian  
 CC tumours in an individual via the detection and measurement of the  
 CC expression of ovarian tumor marker genes (ABA83081-ABA83122, ABA83180,  
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
 CC ABA83181 and ABA83183). The methods of the invention are useful for  
 CC detecting an ovarian tumour in a patient, for identifying an individual  
 CC at increased risk for developing ovarian cancer, in prognostic tests for  
 CC assessing the relative severity of ovarian cancer, in tests for

CC monitoring a patient in remission from ovarian cancer and in tests for  
 CC monitoring disease status in a patient being treated for ovarian cancer.  
 CC The methods can additionally be used to identify a particular tumour as  
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
 CC tumour. The ovarian tumour marker genes of the invention were identified  
 CC using SAGE (serial analysis of gene expression) and were found to be  
 CC overexpressed in a broad variety of ovarian epithelial tumour cells  
 CC relative to normal ovarian epithelial cells. The marker genes are  
 CC implicated in immune response pathways, in the regulation of cell  
 CC proliferation and in protein folding, and many of these are membrane-  
 CC localised or secreted. In addition to their use as diagnostic and  
 CC prognostic markers, the ovarian tumour marker genes or their encoded  
 CC proteins may be used as therapeutic targets for the treatment and  
 CC prevention of ovarian cancer. Sequences ABB50257-ABB50289 represent  
 CC proteins encoded by ovarian tumour marker genes of the invention.

XX SQ Sequence 1466 AA;

Query Match 65.2%; Score 60; DB 22; Length 1466;

Best Local Similarity 78.6%; Pred. No. 3;

Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPGWPCA 14

Db 738 GPKGDKGPGGPGCA 751

RESULT 30

AAE02533 standard; Protein; 1466 AA.

XX AAE02533;

DT 10-AUG-2001 (first entry)

DE Bovine alpha1(III) collagen #1.

KW Bovine; alpha1(III) collagen; gelatin; cytosol; viral infection;  
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
 KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
 KW rheumatoid arthritis; beverage; photographic application.

XX Bos sp.

PN WO200134647-A2.

PD 17-MAY-2001.

PF 10-NOV-2000; 2000MO-US30792.

PR 12-NOV-1999; 99US-0439058.

PR 10-NOV-2000; 2000US-0439058.

PA (FIBR-) FIBROGEN INC.

PI Bell MP, Neff TB, Polarek JW, Seeley TW;

DR WPI; 2001-335911/35.

DR N-PSDB; AAD06574.

PT Novel isolated and purified bovine or porcine collagens and gelatins  
 PT useful in medical, pharmaceutical, food and cosmetic industries, as  
 PT vaccine, and for treating autoimmune disorders, infections and cancer

PS Example 2; Fig 4; 168pp; English.

CC The present sequence is bovine alpha1(III) collagen. The present  
 CC invention relates to recombinant synthesis of collagens and gelatins

CC derived from animals. Collagen is useful in medical, pharmaceutical,  
 CC food and cosmetic industries. Collagen is an important component of  
 CC arterial sealants, bone grafts, drug delivery system, dermal implants,  
 CC hemostats, and incontinence implants, and for treating autoimmune  
 CC disorders such as rheumatoid arthritis. Collagen is useful in food  
 CC products such as sausage casings, and in cosmetics or facial and skin  
 CC products such as moisturisers. Recombinant gelatin is useful in vaccine  
 CC formulations for treating viral infections, autoimmune diseases and  
 CC cancer. Gelatin is useful in the manufacture or as a component of  
 CC various pharmaceutical and medical devices and products, in food and  
 CC beverage industries, in hair care and skin care products, as a glue or  
 CC adhesive in various manufacturing processes, as a light-sensitive coating  
 CC in various electronic devices, as photoresist base in photolithographic  
 CC processes, in printing and photographic applications, in laboratory  
 CC application, and as a component in various gels used for biochemical and  
 CC electrophoretic analysis, including enzymographic gels.

XX SQ Sequence 1466 AA;

Query Match 65.2%; Score 60; DB 22; Length 1466;

Best Local Similarity 71.4%; Pred. No. 3;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPGWPCA 14

Db 858 GVGKDGKNGPGWPCA 871

Search completed: November 1, 2002, 12:52:33

Job time : 29.5 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 1, 2002, 12:52:42 : Search time 10.5 Seconds

(Without alignments)  
34.894 Million cell updates/sec

Title: US-09-529-691A-1

Sequence: 92  
1 GVAGDKGNPGWPGAP 15Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed.  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	15	3 US-08-394-748A-1	Sequence 1, Appl1
2	92	100.0	15	3 US-08-702-254A-1	Sequence 1, Appl1
3	92	100.0	15	5 PCT-US95-02478-1	Sequence 1, Appl1
4	92	100.0	24	1 US-08-534-342-3	Sequence 3, Appl1
5	92	100.0	24	1 US-08-675-140-3	Sequence 3, Appl1
6	92	100.0	39	1 US-08-534-342-2	Sequence 2, Appl1
7	92	100.0	39	1 US-08-675-140-2	Sequence 2, Appl1
8	66	71.7	39	1 US-08-534-342-5	Sequence 5, Appl1
9	66	71.7	39	1 US-08-534-342-6	Sequence 6, Appl1
10	66	71.7	39	1 US-08-675-140-5	Sequence 5, Appl1
11	66	71.7	39	1 US-08-675-140-6	Sequence 6, Appl1
12	61	66.3	532	1 US-08-494-168-9	Sequence 9, Appl1
13	60	65.2	1057	3 US-08-931-820-4	Sequence 4, Appl1
14	60	65.2	1078	4 US-08-963-825-21	Sequence 21, Appl1
15	60	65.2	1078	4 US-09-570-573-21	Sequence 21, Appl1
16	60	65.2	1078	4 US-09-548-608-21	Sequence 21, Appl1
17	59	64.1	471	2 US-08-399-889-24	Sequence 24, Appl1
18	59	64.1	471	3 US-09-167-364-24	Sequence 24, Appl1
19	59	64.1	471	4 US-09-439-897-2	Sequence 2, Appl1
20	56	60.9	489	2 US-08-794-795-7	Sequence 7, Appl1
21	56	60.9	489	4 US-09-249-200-7	Sequence 7, Appl1
22	56	60.9	518	3 US-08-392-367B-2	Sequence 2, Appl1
23	56	60.9	518	3 US-08-893-467A-2	Sequence 2, Appl1
24	56	60.9	546	1 US-08-494-168-10	Sequence 10, Appl1
25	55	59.8	171	3 US-09-011-735-2	Sequence 2, Appl1
26	55	59.8	351	3 US-09-011-735-1	Sequence 1, Appl1
27	55	59.8	1694	1 US-08-494-168-27	Sequence 2, Appl1

28	54	58.7	228	4 US-09-219-849-38	Sequence 38, Appl1
29	53	57.6	495	2 US-08-794-795-2	Sequence 2, Appl1
30	53	57.6	495	4 US-09-249-200-2	Sequence 2, Appl1
31	53	57.6	520	2 US-08-794-795-6	Sequence 6, Appl1
32	53	57.6	520	4 US-09-249-200-6	Sequence 6, Appl1
33	53	57.6	549	1 US-08-494-168-8	Sequence 8, Appl1
34	53	57.6	557	3 US-09-320-095-10	Sequence 10, Appl1
35	53	57.6	557	4 US-09-523-887-10	Sequence 10, Appl1
36	52	56.5	279	4 US-09-010-999-2	Sequence 2, Appl1
37	52	56.5	1057	3 US-08-931-820-1	Sequence 1, Appl1
38	52	56.5	1060	3 US-08-931-820-3	Sequence 3, Appl1
39	52	56.5	1096	4 US-09-346-237-6	Sequence 6, Appl1
40	52	56.5	1418	3 US-08-963-825-20	Sequence 20, Appl1
41	52	56.5	1418	4 US-09-010-999-1	Sequence 1, Appl1
42	52	56.5	1418	4 US-09-570-573-20	Sequence 20, Appl1
43	52	56.5	1418	4 US-09-548-608-20	Sequence 20, Appl1
44	52	56.5	1442	2 US-08-316-650-12	Sequence 12, Appl1
45	52	56.5	1442	5 PCT-US95-02251-12	Sequence 12, Appl1

## ALIGNMENTS

RESULT 1  
US-08-394-748A-1  
Sequence 1, Application US/08394748A  
Patent No. 6013628  
GENERAL INFORMATION:  
APPLICANT: Skubitz, Amy P.N.  
APPLICANT: Furcht Leo T.  
APPLICANT: Bailes, Mark  
APPLICANT: Gregerson, Dale S.  
APPLICANT: Agarwal, Anita  
APPLICANT: Wright, Martha M.  
APPLICANT: Murali, Shobana  
TITLE OF INVENTION: Method for Treating Conditions of the Eye  
TITLE OF INVENTION: Using Polypeptides  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 6013628west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,748A  
FILING DATE: 27-FEB-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/203,458  
FILING DATE: 28-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Carter, Charles G.  
REGISTRATION NUMBER: 35,093  
REFERENCE/DOCKET NUMBER: 600,307US01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
ORIGINAL SOURCE:  
STRAIN: IV-H1  
US-08-394-748A-1

Query Match 100.0%; Score 92; DB 3; Length 15;  
Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGDKGNPGWPGAP 15  
DB 1 GVGDKGNPGWPGAP 15

## RESULT 2

US-08-702-254A-1

; Sequence 1, Application US/08702254A  
; Patent No. 6096863

; GENERAL INFORMATION:

; APPLICANT: FIELDS, GREGG B.

; TITLE OF INVENTION: SELF-ASSEMBLING AMPHIPHILES FOR

; TITLE OF INVENTION: CONSTRUCTION OF PEPTIDE SECONDARY STRUCTURES

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MUEITING, RAASCH &amp; GERHARDT, P.A.

; STREET: 119 NORTH FOURTH STREET

; CITY: MINNEAPOLIS

; STATE: MINNESOTA

; COUNTRY: USA

; ZIP: 55401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/702.254A

; FILING DATE: 23-AUG-1996

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: MUEITING, ANN M.

; REGISTRATION NUMBER: 33,977

; REFERENCE/DOCKET NUMBER: 110.00310101

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-305-1217

; TELEFAX: 612-305-1228

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-702-254A-1

; Query Match 100.0%; Score 92; DB 3; Length 15;

; Best Local Similarity 100.0%; Pred. No. 2.1e-07;

; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGDKGNPGWPGAP 15

DB 1 GVGDKGNPGWPGAP 15

## RESULT 3

PCT-US95-02478-1

; Sequence 1, Application PC/TUS9502478

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Method for Treating Conditions

; TITLE OF INVENTION: Of the Eye Using Polypeptides

; NUMBER OF SEQUENCES: 16

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/02478

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/203,458

; FILING DATE: 28-FEB-1994

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 15 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; ORIGINAL SOURCE:

; STRAIN: IV-H1

; PCT-US95-02478-1

; Query Match 100.0%; Score 92; DB 5; Length 15;

; Best Local Similarity 100.0%; Pred. No. 2.1e-07;

; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGDKGNPGWPGAP 15

DB 1 GVGDKGNPGWPGAP 15

## RESULT 4

US-08-534-342-3

; Sequence 3, Application US/08534342

; Patent No. 5576419

; GENERAL INFORMATION:

; APPLICANT: Fields, Gregg B.

; TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant &amp; Gould

; STREET: 3100 No. 5576419west Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/534,342

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/085,633

; FILING DATE: 30-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Mueitling, Ann M.

; REGISTRATION NUMBER: 33,977

; REFERENCE/DOCKET NUMBER: 600.253-US-01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-332-5300

; TELEFAX: 612-332-9081

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 24 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

; US-08-534-342-3

; Query Match 100.0%; Score 92; DB 1; Length 24;

; Best Local Similarity 100.0%; Pred. No. 3.3e-07;

; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GVGDKGNPGWPGAP 15

DB 1 GVGDKGNPGWPGAP 15

Db 10 GVKGDKGNPGMPCAP 24

RESULT 5  
US-08-675-140-3; Sequence 3, Application US/08675140  
; Patent No. 5726243

; GENERAL INFORMATION:

; APPLICANT: Fields, Gregg B.

; TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant &amp; Gould

; STREET: 3100 No. 5726243west Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/675,140

; FILING DATE: 03-JUL-1996

; CLASSIFICATION: 525

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/085,633

; FILING DATE: 30-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Mueller, Ann M.

; REGISTRATION NUMBER: 33,977

; REFERENCE/DOCKET NUMBER: 600,253-US-01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-332-5300

; TELEFAX: 612-332-9081

; INFORMATION FOR SEQ. ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 24 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

US-08-675-140-3

Query Match 100.0%; Score 92; DB 1; Length 24;  
Best Local Similarity 100.0%; Pred. No. 3.3e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 GVKGDKGNPGMPCAP 15  
Db 10 GVKGDKGNPGMPCAP 24RESULT 6  
US-08-534-342-2

; Sequence 2, Application US/08534342

; Patent No. 5576419

; GENERAL INFORMATION:

; APPLICANT: Fields, Gregg B.

; TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant &amp; Gould

; STREET: 3100 No. 5576419west Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/534,342

FILING DATE: 530

CLASSIFICATION: 530

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/085,633

FILING DATE: 30-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Mueller, Ann M.

REGISTRATION NUMBER: 33,977

REFERENCE/DOCKET NUMBER: 600,253-US-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

INFORMATION FOR SEQ. ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-534-342-2

Query Match 100.0%; Score 92; DB 1; Length 39;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 GVKGDKGNPGMPCAP 15  
Db 25 GVKGDKGNPGMPCAP 39RESULT 7  
US-08-675-140-2

; Sequence 2, Application US/08675140

; Patent No. 5726243

; GENERAL INFORMATION:

; APPLICANT: Fields, Gregg B.

; TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Merchant &amp; Gould

; STREET: 3100 No. 5726243west Center

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/675,140

; FILING DATE: 03-JUL-1996

; CLASSIFICATION: 525

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/085,633

; FILING DATE: 30-JUN-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Mueller, Ann M.

; REGISTRATION NUMBER: 33,977

; REFERENCE/DOCKET NUMBER: 600,253-US-01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-332-5300

; TELEFAX: 612-332-9081

; INFORMATION FOR SEQ. ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 39 amino acids

; TYPE: amino acid

; TOPOLOGY: unknown

US-08-675-140-2

Query Match 100.0%; Score 92; DB 1; Length 39;  
Best Local Similarity 100.0%; Pred. No. 5.6e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GYKGDKNPWPAP 15  
|||||  
Db 25 GYKGDKNPWPAP 39

RESULT 8  
US-08-534-342-5  
Sequence 5, Application US/08534342  
Patent No. 5576419  
GENERAL INFORMATION:  
APPLICANT: Fields, Gregg B.  
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merchant & Gould  
STREET: 3100 No. 5576419west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/534,342  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/085,633  
FILING DATE: 30-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueeling, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 600.253-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
US-08-534-342-5

Query Match 71.7%; Score 66; DB 1; Length 39;  
Best Local Similarity 73.3%; Pred. No. 0.003;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GYKGDKNPWPAP 15  
|||||  
Db 25 GYKGDKNPWPAP 39

RESULT 9  
US-08-534-342-6  
Sequence 6, Application US/08534342  
Patent No. 5576419  
GENERAL INFORMATION:  
APPLICANT: Fields, Gregg B.  
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merchant & Gould  
STREET: 3100 No. 5576419west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/534,342  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/085,633  
FILING DATE: 30-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueeling, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 600.253-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
US-08-534-342-6

Query Match 71.7%; Score 66; DB 1; Length 39;  
Best Local Similarity 73.3%; Pred. No. 0.003;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GYKGDKNPWPAP 15  
|||||  
Db 25 GYKGDKNPWPAP 39

RESULT 10  
US-08-675-140-5  
Sequence 5, Application US/08675140  
Patent No. 5726243  
GENERAL INFORMATION:  
APPLICANT: Fields, Gregg B.  
TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESS: Merchant & Gould  
STREET: 3100 No. 5726243west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/675,140  
FILING DATE: 03-JUL-1996  
CLASSIFICATION: 525  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/085,633  
FILING DATE: 30-JUN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueeling, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 600.253-US-01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown





LOCATION: 1055  
OTHER INFORMATION: /label= Modified  
OTHER INFORMATION: /note= "Ala may be Pro"

US-08-931-820-4

Query Match 65.2%; Score 60; DB 3; Length 1057;  
Best Local Similarity 71.4%; Pred. No. 0.67;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDGNGPMPGA 14  
| | | | : | : | | | | |  
DB 710 GVKGERGSPGGPGA 723

RESULT 14  
US-08-963-825-21  
Sequence 21, Application US/08963825  
Patent No. 6110689

GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,825  
FILING DATE:  
CLASSIFICATION: 436

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21

Query Match 65.2%; Score 60; DB 3; Length 1078;  
Best Local Similarity 71.4%; Pred. No. 0.69;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDGNGPMPGA 14  
| | | | : | : | | | | |  
DB 711 GVKGERGSPGGPGA 724

RESULT 15  
US-09-570-573-21  
Sequence 21, Application US/09570573  
Patent No. 6342361

GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-570-573-21

Query Match 65.2%; Score 60; DB 4; Length 1078;  
Best Local Similarity 71.4%; Pred. No. 0.69;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDGNGPMPGA 14  
| | | | : | : | | | | |  
DB 711 GVKGERGSPGGPGA 724

RESULT 16  
US-09-548-608-21  
Sequence 21, Application US/09548608  
Patent No. 6355442

GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/548,608
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogosis, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (III)
US-09-548-608-21

Query Match 65.2%; Score 60; DB 4; Length 1078;
Best Local Similarity 71.4%; Pred. No. 0.69;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDGKGNPGMPCA 14
Db 711 GVKGRSGSGPGGA 724

RESULT 17
US-08-399-889-24
Sequence 24, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-08-399-889-24

Query Match 64.1%; Score 59; DB 2; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.4;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 GVKDGKGNPGMPCAP 15
Db 116 GVKGRKNSGSGPGPP 130

RESULT 18
US-09-167-364-24
Sequence 24, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-09-167-364-24

Query Match 64.1%; Score 59; DB 3; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.4;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKDGKGNPGMPCAP 15
Db 116 GVKGRKNSGSGPGPP 130

RESULT 19
US-09-439-897-2
Sequence 2, Application US/09439897
Patent No. 6277558
GENERAL INFORMATION:
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 95-1263-C
CURRENT APPLICATION NUMBER: US/09/439,897
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 65
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 471
TYPE: PRT
ORGANISM: Bos taurus
US-09-439-897-2

Query Match 64.1%; Score 59; DB 4; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.4;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKDGKGNPGMPCAP 15
Db 116 GVKGRKNSGSGPGPP 130

RESULT 20
US-08-794-795-7
Sequence 7, Application US/08794795
Patent No. 5916766
GENERAL INFORMATION:
APPLICANT: Eshourlagy, Nabih
APPLICANT: Adamou, John
APPLICANT: Gross, Mitchell
APPLICANT: Lyeko, Paul
```

;; TITLE OF INVENTION: Human Macro Scavenger Rec  
;; TITLE OF INVENTION: eptor  
;; NUMBER OF SEQUENCES: 9  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Smithline Beecham Corporation  
;; STREET: 709 Swedeland Road  
;; CITY: King of Prussia  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19406  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/794,795  
;; FILING DATE: 04-FEB-1997  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: ATG50009P  
;; FILING DATE: 22-MAY-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Han, William T  
;; REGISTRATION NUMBER: 34,344  
;; REFERENCE/DOCKET NUMBER: ATG50009  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-270-5219  
;; TELEFAX: 610-270-4026  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 489 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-794-795-7  
;  
Query Match 60.9%; Score 56; DB 2; Length 489;  
Best Local Similarity 66.7%; Pred. No. 1.1;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
;  
QY 1 GVKGDKNPGWPGAP 15  
Db 265 GVKGDKNPGWPGAP 279  
;  
RESULT 21  
US-09-249-200-7  
;; Sequence 7, Application US/09249200  
;; Patent No. 6197931  
;; GENERAL INFORMATION:  
;; APPLICANT: ELSHOURBAGY, NABIL  
;; APPLICANT: ADAMOU, JOHN  
;; APPLICANT: GROSS, MITCHELL  
;; APPLICANT: LYSKO, PAUL  
;; TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR  
;; NUMBER OF SEQUENCES: 9  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Ratner & Prestlia  
;; STREET: P.O. Box 980  
;; CITY: Valley Forge  
;; STATE: PA  
;; COUNTRY: USA  
;; ZIP: 19482  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: FASTSEQ for Windows Version 2.0  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/249,200

;; FILING DATE: 12-FEB-1999  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 08/794,795  
;; FILING DATE: 04-FEB-1997  
;; APPLICATION NUMBER: 60/017,699  
;; FILING DATE: 23-MAY-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Prestlia, Paul F  
;; REGISTRATION NUMBER: 23,031  
;; REFERENCE/DOCKET NUMBER: ATG-50009-1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 610-407-0700  
;; TELEFAX: 610-407-0700  
;; TELEX: 846169  
;; INFORMATION FOR SEQ ID NO: 7:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 489 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-09-249-200-7  
;  
Query Match 60.9%; Score 56; DB 4; Length 489;  
Best Local Similarity 66.7%; Pred. No. 1.1;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
;  
QY 1 GVKGDKNPGWPGAP 15  
Db 265 GVKGDKNPGWPGAP 279  
;  
RESULT 22  
US-08-392-367B-2  
;; Sequence 2, Application US/08392367B  
;; Patent No. 5691197  
;; GENERAL INFORMATION:  
;; APPLICANT: Trygvason, Karl  
;; APPLICANT: Elomaa, Outi  
;; APPLICANT: Kangas, Maarit  
;; TITLE OF INVENTION: An Insolated DNA Sequence For a  
;; Patent No. 5691197  
;; TITLE OF INVENTION: No. 5691197e1 Macrophage Receptor with  
;; TITLE OF INVENTION: a Collagenous Domain and the  
;; TITLE OF INVENTION: Polypeptide Chain Encoded by  
;; NUMBER OF SEQUENCES: 2  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
;; ADDRESSEE: Minnich & McKee  
;; STREET: 1100 Superior Avenue  
;; STREET: Suite 700  
;; CITY: Cleveland  
;; STATE: Ohio  
;; COUNTRY: U.S.A.  
;; ZIP: 44114-2518  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch,  
;; MEDIUM TYPE: 720 Kb storable  
;; COMPUTER: IBM PS/2, Model 35 SX  
;; OPERATING SYSTEM: DOS 5.0  
;; SOFTWARE: Word Perfect 5.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/392,367B  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Minnich, Richard J.  
;; REGISTRATION NUMBER: 24,175  
;; REFERENCE/DOCKET NUMBER: TRV 2 009  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (216) 861-5582

TELEFAX: (216) 241-1666  
TELEX: (216) 980162  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 amino acids  
TYPE: amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-392-367B-2

Query Match 60.9%; Score 56; DB 1; Length 518;  
Best Local Similarity 66.7%; Pred. No. 1.2;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVKDGNPGMPGAP 15  
Db 294 GVKDGNPGMPGAP 308

RESULT 23  
US-08-893-467A-2  
Sequence 2, Application US/08893467A  
Patent No. 6063901  
GENERAL INFORMATION:  
APPLICANT: Trygvason, Karl  
APPLICANT: Elomaa, Outi  
APPLICANT: Kangas, Maarit  
TITLE OF INVENTION: An Insolated DNA Sequence For a  
Patent No. 6063901  
TITLE OF INVENTION: No. 6063901el Macrophage Receptor with  
TITLE OF INVENTION: a Collagenous Domain and the  
TITLE OF INVENTION: Polypeptide Chain Encoded by  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
ADDRESSEE: Minnich & McKee  
STREET: 1100 Superior Avenue  
STREET: Suite 700  
CITY: Cleveland  
STATE: Ohio  
COUNTRY: U.S.A.  
ZIP: 44114-2518  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch,  
COMPUTER: IBM PS/2, Model 35 SX  
OPERATING SYSTEM: DOS 5.0  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/893,467A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Minnich, Richard J.  
REGISTRATION NUMBER: 24,175  
REFERENCE/DOCKET NUMBER: TRV 2 009  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (216) 861-5582  
TELEFAX: (216) 241-1666  
TELEX: (216) 980162  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 518 amino acids  
TYPE: amino acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-893-467A-2

Query Match 60.9%; Score 56; DB 3; Length 518;  
Best Local Similarity 66.7%; Pred. No. 1.2;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVKDGNPGMPGAP 15  
Db 294 GVKDGNPGMPGAP 308

RESULT 24  
US-08-494-168-10  
Sequence 10, Application US/08494168  
Patent No. 5731192  
GENERAL INFORMATION:  
APPLICANT: Reeders, Stephen T.  
APPLICANT: Zhou, Jing  
TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method  
TITLE OF INVENTION: of Detecting Collagen Deficiency  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/494,168  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/112,465  
FILING DATE: 27-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Bernhard D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 40397/104/BABR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-494-168-10

Query Match 60.9%; Score 56; DB 1; Length 546;  
Best Local Similarity 75.0%; Pred. No. 1.3;  
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 4 GVKDGNPGMPGAP 15  
Db 168 GVKDGNPGMPGAP 179

RESULT 25  
US-09-011-735-2  
Sequence 2, Application US/09011735B  
Patent No. 6110708  
GENERAL INFORMATION:  
APPLICANT: Wakamiya, No. 6110708utaka  
TITLE OF INVENTION: Recombinant Conjugulin and Producing Method Thereof  
FILE REFERENCE: 19036/34548  
CURRENT APPLICATION NUMBER: US/09/011,735B  
CURRENT FILING DATE: 1998-05-22  
EARLIER APPLICATION NUMBER: JP 7-209698  
EARLIER FILING DATE: 1995-08-17  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 2  
LENGTH: 171



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/794,795
FILING DATE: 04-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: ATG50009P
FILING DATE: 22-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-794-795-2

Query Match          57.6%; Score 53; DB 2; Length 495;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GVKDKGNPGMGAP 15
Db 270 GAKGDGQPGGLGVP 284

RESULT 30
US-09-249-200-2
Sequence 2, Application US/09249200
Patent No. 6197931
GENERAL INFORMATION:
APPLICANT: ELSHOURBAGY, NABIL
APPLICANT: ADAMOU, JOHN
APPLICANT: GROSS, MITCHELL
APPLICANT: LYSKO, PAUL
TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249,200
FILING DATE: 12-FEB-1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
```

```

APPLICATION NUMBER: 08/794,795
FILING DATE: 04-FEB-1997
APPLICATION NUMBER: 60/017,699
FILING DATE: 23-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Prestia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: ATG-50009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-249-200-2

Query Match          57.6%; Score 53; DB 4; Length 495;
Best Local Similarity 60.0%; Pred. No. 3.1;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GVKDKGNPGMGAP 15
Db 270 GAKGDGQPGGLGVP 284
```

Search completed: November 1, 2002, 12:54:44  
Job time : 11.5 secs

5 PAGE BLANK (USPTO)



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:38:32 ; Search time 13.5 Seconds  
(without alignments)  
106.766 Million cell updates/sec

Title: US-09-529-691A-1

Perfect score: 92  
Sequence: 1 GVKCDKNGNPGMGAP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database :

1: PIR\_71:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	1669	1 CGH04B	collagen alpha 1(I
2	78	84.8	1669	1 CGH04B	collagen alpha 1(I
3	72	78.3	356	1 S16907	collagen alpha 1(I
4	69	75.0	1691	1 S22917	collagen alpha 5(I
5	66	71.7	754	2 A55267	collagen alpha 5(I
6	65	70.7	1707	2 A33526	collagen alpha 2(I
7	64	69.6	1712	1 CGH02B	collagen alpha 2(I
8	63	68.5	677	2 S23296	collagen alpha 2(I
9	63	68.5	2944	2 A54849	collagen alpha 1(V
10	62	67.4	290	2 A32249	collagen - sea urc
11	62	67.4	1414	1 S23809	collagen alpha 2(I
12	62	67.4	1670	1 CGH03B	collagen alpha 3(I
13	61	66.3	340	1 B46345	gene V protein - p
14	61	66.3	1603	2 S23810	collagen alpha 1(I
15	61	66.3	1690	1 CGH01B	collagen alpha 1(I
16	61	66.3	1763	2 S16366	collagen alpha 4(I
17	60	65.2	310	2 S15096	collagen alpha 1(I
18	60	65.2	623	2 A45137	collagen alpha 4(I
19	60	65.2	636	2 S41067	collagen alpha 1(I
20	60	65.2	886	2 I50694	collagen alpha 1(I
21	60	65.2	1049	1 CGH07S	collagen alpha 1(I
22	60	65.2	1466	1 CGH07L	collagen alpha 1(I
23	60	65.2	1532	2 A61262	collagen alpha 1(I
24	60	65.2	1761	2 T13990	collagen type IV a
25	59.5	64.7	1752	2 A45607	collagen alpha 3(I
26	59	64.1	369	2 S33603	surfactant protein
27	59	64.1	471	2 A39024	collagen alpha 3(I
28	59	64.1	1433	2 A46053	bullous pemphigoid
29	59	64.1	1744	2 S40991	collagen alpha 1(I

30	58	63.0	285	2 S60598	collagen alpha cha
31	58	63.0	464	2 S59513	collagen II Al pro
32	58	63.0	1758	2 T29350	hypothetical prote
33	58	63.0	1759	2 T29351	collagen alpha 2(I
34	57	62.0	298	2 TC1448	collagen col-34 -
35	57	62.0	650	2 T22002	hypothetical prote
36	57	62.0	920	2 A45748	collagen alpha 1(V
37	57	62.0	1549	2 I48103	type VII collagen
38	56	60.9	299	2 T24833	hypothetical prote
39	56	60.9	371	2 E88633	protein F56B3.111
40	56	60.9	380	2 A46295	collagen I - marin
41	56	60.9	518	2 A55840	macrophage bacteri
42	56	60.9	775	2 A61228	collagen alpha 2(I
43	56	60.3	1142	2 JX0369	collagen alpha 1(X
44	55.5	60.3	350	2 T73338	macrophage scaveng
45	55.5	60.3	458	2 B44407	macrophage scaveng

#### ALIGNMENTS

RESULT 1  
CGH04B  
collagen alpha 1(IV) chain precursor - human  
N:Alternate names: procollagen alpha 1(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 28-May-1986 #sequence, revision 31-Dec-1992 #text, change 07-Dec-1999  
C:Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863;  
R:Solinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Trygvaeson, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV c  
A:Reference number: S16876; M0ID:89340433  
A:Accession: S16876  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1669 <SO11>  
A:Cross-references: EMBL:J04217; GB:J05039; NID:G180800; PIDN:AA53098.1; PID:G180803  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R:Solinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Trygvaeson, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen  
A:Reference number: S02738; M0ID:8930632  
A:Accession: S02738  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-6, 'L', '8-28 <POE>  
A:Cross-references: EMBL:X12784; NID:G30072  
R:Bratzel, D.; Oberbauer, I.; Dierlinger, H.; Babel, W.; Glanville, R.W.; Deutzmann, Eur. J. Biochem. 168, 529-536, 1987  
A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement  
A:Reference number: S00048; M0ID:88029471  
A:Accession: S00048  
A:Molecule type: mRNA  
A:Residues: 1-318, 'A', '320-944 <BRA1>  
A:Cross-references: EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID:G30067  
A:Accession: S25826  
A:Molecule type: protein  
A:Residues: 271-318, 'A', '320-554 <BRA2>  
R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Ristell, J.; Kuehn, K.  
Eur. J. Biochem. 152, 213-219, 1985  
A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region  
A:Reference number: A23115; M0ID:86004708  
A:Accession: A23115  
A:Molecule type: protein  
A:Residues: 28-236, 'KE', '239-240, 'K', '242-243 <GLA>



A:Residues: 1-185,'L',187-318,'S',320-368,'L',370-402,'F',404-480,'L',482-492,'H',494-71  
A:Cross-references: EMBL:X06777  
R.Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.  
J. Biol. Chem. 263, 8706-8709, 1988  
A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
A:Reference number: A28066; MUID:88243724  
A:Accession: A28066  
A:Molecule type: mRNA  
A:Residues: 1-129 <KII>  
A:Cross-references: EMBL:J03758; NID:g192669; PIDN:AAA37439.1; PID:g192670  
R.Oberdaeme, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogell, G.; Voss,  
Eur. J. Biochem. 147, 217-224, 1985  
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
A:Reference number: A02864; MUID:85127033  
A:Accession: A02864  
A:Molecule type: mRNA  
A:Residues: 1276-1669 <OBE>  
A:Cross-references: EMBL:X02201; NID:g50233; PIDN:CAA26132.1; PID:g1333876  
R.Math, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogell, G.  
Gene 43, 301-304, 1986  
A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox  
A:Reference number: A25636; MUID:86301866  
A:Accession: A25636  
A:Molecule type: mRNA  
A:Residues: 1149-1396,'S',1398-1424 <NAT>  
A:Cross-references: EMBL:M14042; NID:g192286; PIDN:AAA37342.1; PID:g192287  
A:Note: The authors translated the codon CAG for residue 1374 as Arg  
R.Kurinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj  
J. Biol. Chem. 262, 8496-8499, 1987  
A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
A:Reference number: A94680; MUID:87250460  
A:Accession: A29301  
A:Molecule type: mRNA  
A:Residues: 1441-1669 <KUR>  
A:Cross-references: EMBL:M15832; NID:g192282; PIDN:AAA37340.1; PID:g387115  
R.Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y.  
J. Biol. Chem. 263, 12310-12314, 1988  
A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc  
A:Reference number: S19079; MUID:86315019  
A:Accession: S19079  
A:Molecule type: DNA  
A:Residues: 1-28 <K12>  
A:Cross-references: EMBL:J03944; NID:g192673; PIDN:AAA37442.1; PID:g466503  
R.Kayes, P.; Wood, L.; Theriault, N.; Kurinen, M.; Vogell, G.  
J. Biol. Chem. 263, 19274-19277, 1988  
A:Title: Head-to-head arrangement of murine type IV collagen genes.  
A:Reference number: A92702; MUID:89066738  
A:Accession: A32003  
A:Molecule type: DNA  
A:Residues: 1-28 <KAY>  
A:Cross-references: EMBL:J04448; NID:g192666; PIDN:AAA37437.1; PID:g450449  
R.Burbelo, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988  
A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
A:Reference number: A94220; MUID:89071759  
A:Accession: A31766  
A:Molecule type: DNA  
A:Residues: 1-28 <BUR>  
A:Cross-references: EMBL:M23333; NID:g340878; PIDN:AAA51625.1; PID:g535668  
R.Sakurai, Y.; Sullivan, M.; Yamada, Y.  
J. Biol. Chem. 261, 6654-6657, 1986  
A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
A:Reference number: S19094; MUID:86196099  
A:Accession: S19094  
A:Molecule type: DNA  
A:Residues: 1110-1135;1189-1316;1342-1383;1418-1487 <SAR>  
A:Cross-references: EMBL:M13027  
R.Schuppan, D.; Timpl, R.; Glanville, R.W.  
FEBS Lett. 115, 297-300, 1980  
A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane colla  
A:Reference number: S16909; MUID:80246483  
A:Accession: S16909  
A:Molecule type: protein

A:Residues: 940-946,'G',948-949,'G',951-955,'G',957;1213-1228,'X',1230-1234,'P',1236-  
R.Schuppan, D.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 123, 505-512, 1982  
A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a  
A:Reference number: A25991; MUID:82186723  
A:Accession: A25991  
A:Molecule type: protein  
A:Residues: 940-946,'X',948-949,'X',951-955,'X',957-964,'X',966-991,'X',993-1003,'X',  
61,'X',1063-1065,'X',1067-1080,'X',1082-1083,'X',1085-1106,'X',1108-1115,'DE',1118-11  
A:Accession: B25991  
A:Molecule type: protein  
A:Residues: 1173-1181,'X',1183-1184,'X',1186-1187,'X',1189-1205,'O',1207,'XE',1210-12  
3,'SP',1266,'IT',1269,'SK',1272,'DM',1275,'L',1277-1282;1316-1318,'X',1320-1327,'X',1  
R.Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 139, 401-410, 1984  
A:Title: Subunit structure and assembly of the globular domain of basement-membrane c  
A:Reference number: S17801; MUID:84132058  
A:Accession: S17801  
A:Molecule type: protein  
A:Residues: 1435-1443 <WEB>  
C:Genetics:  
A:introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3  
A:Note: The list of introns may be incomplete  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coll.; duplication; extracellular  
F:1-27/Domain: signal sequence #status predicted <SIG>  
F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <NAT>  
F:28-162/Domain: 75 <7SD>  
F:163-1440/Domain: collagenous, triple helix <COL>  
F:597-599/Region: cell attachment (R-G-D) motif  
F:781-783/Region: cell attachment (R-G-D) motif  
F:917-919/Region: cell attachment (R-G-D) motif  
F:968-970/Region: cell attachment (R-G-D) motif  
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F:1553-1669/Region: duplication  
F:11,36,39,41,434,467,470/Duplicate bonds: Interchain #status predicted  
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:917,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,109  
92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxy  
F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:1505-1511,1616-1622/Duplicate bonds: #status predicted

Query Match 84.8%; Score 78; DB 1; Length 1669;  
Best local similarity 86.7%; Pred. No. 0.0018;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GYKDKGNPGWPGAP 15  
DB 1263 GPKDKGNQWPGAP 1277

RESULT 3  
S16907  
collagen alpha 1(IV) chain - bovine (fragments)  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 04-Dec-1992 #sequence, revision 05-Apr-1995 #text, change 19-Oct-1995  
C:Accession: A39474; S16907; S18432; A39419  
R:Gunwar, S.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 266, 14088-14094, 1991  
A:Title: Properties of the collagenous domain of the alpha3(IV) chain, the Goodpastur  
le helical structure and noncollagenous domain.  
A:Reference number: A39474; MUID:91310700  
A:Accession: A39474  
A:Molecule type: protein  
A:Residues: 1-9 <GUZ>  
R.Schuppan, D.; Glanville, R.W.; Timpl, R.; Dixit, S.N.; Kang, A.H.  
Biochem. J. 220, 227-233, 1984  
A:Title: Sequence comparison of peptic-resistant segments of basement-membrane colla  
A:Reference number: S16907; MUID:84256630  
A:Accession: S16907  
A:Molecule type: protein

A:Residues: 4-8,'E',10-66;69-158;159-321;322-349 <SC2>  
R:Butkowsk, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7874-7877, 1987  
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membrane  
A:Reference number: S18432; MUID:87222419  
A:Accession: S18432  
A:Molecule type: protein  
A:Residues: 337-347,'FL',350-353 <BU2>  
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe  
J. Biol. Chem. 266, 15318-15324, 1991  
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll  
A:Reference number: A39419; MUID:91332055  
A:Accession: A39419  
A:Molecule type: protein  
A:Residues: 337-347,'FL',350-356 <GU3>  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coll; disulfide bond; extracellular  
F:31,34,37,46,61,69,78,84,87,102,110,122,125,137,140,143,149,155,158,161,164,185,188,195  
proline (Pro) #status experimental

Query Match 78.3%; Score 72; DB 2; Length 356;  
Best Local Similarity 66.7%; Pred. No. 0.003;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKNPGMPAD 15  
I:|||||||  
Db 248 GLXGERGNPGMPGP 262

## RESULT 4

collagen alpha 5(IV) chain precursor, renal splice form - human  
N:Alternate names: procollagen alpha 5(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 30-Sep-1993 #sequence,revision 27-Feb-1997 #text,change 21-Jul-2000  
C:Accession: S22917; A34363; A57079; A37122; T54317; A34850; S18850; I56971; I76598; A35  
R:Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 267, 12475-12481, 1992  
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident  
n Alport syndrome patient.  
A:Reference number: S22917; MUID:92316923  
A:Accession: S22917  
A:Molecule type: mRNA  
A:Residues: 1-967 <ZHO>  
A:Cross-references: GB:M90464; NID:q180826; PIDN:AAA52046.1; PID:q553234  
R:Zhou, J.; Leinonen, A.; Tryggvason, K.  
J. Biol. Chem. 269, 6608-6614, 1994  
A:Title: Structure of the human type IV collagen COL4A5 gene.  
A:Reference number: A54365; MUID:94165049  
A:Accession: A54365  
A:Molecule type: DNA  
A:Residues: 1-922 <ZHZ>  
A:Cross-references: GB:U04470; NID:q463378; GB:U04520; NID:q643428; PIDN:AA027816.1; PID  
R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurilla, P.; de Paeppe, A.; Tryggvas  
Science 261, 1167-1169, 1993  
A:Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm  
A:Reference number: A57079; MUID:93361972  
A:Accession: A57079  
A:Molecule type: DNA  
A:Residues: 1-27 <ZHA>  
A:Cross-references: GB:Z37153; NID:q587203; PIDN:CAA85512.1; PID:q587204  
R:Phlajantemi, T.; Pohjolaenen, E.R.; Myers, J.C.  
J. Biol. Chem. 265, 13758-13766, 1990  
A:Title: Complete primary structure of the triple-helical region and the carboxyl-termi  
A:Reference number: A37122; MUID:90337990  
A:Accession: A37122  
A:Molecule type: mRNA  
A:Residues: 84-433,'SS',442-624,'TALQ',629-666,'FR',669-887,'R',889-1264,1271-1691 <PIH>  
A:Cross-references: GB:U05558; EMBL:M58526; NID:q1314209  
A:Note: submitted to the EMBL data library, February 1991  
A:Note: the authors translated the codon GCC for residue 115 as Val  
R:Renieri, A.; Seri, M.; Myers, J.C.; Phlajantemi, T.; Massella, L.; Rizzoni, G.; De Ma

Hum. Mol. Genet. 1, 127-129, 1992  
A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid  
A:Reference number: I54317; MUID:93244772  
A:Accession: I54317  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 313-324,'E',326-330 <REN>  
A:Cross-references: GB:S5934; NID:q299946; PIDN:AA013909.1; PID:q4261609  
R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.  
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990  
A:Title: Identification of a distinct type IV collagen alpha chain with restricted ki  
A:Reference number: A34850; MUID:90160375  
A:Accession: A34850  
A:Molecule type: mRNA  
A:Residues: 914-1264,1271-1691 <HOS>  
A:Cross-references: EMBL:M3115; NID:q180824; PIDN:AAA52045.1; PID:q180825  
R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.  
Genomics 9, 1-9, 1991  
A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene t  
A:Reference number: A37969; MUID:91169491  
A:Accession: S18850  
A:Molecule type: DNA  
A:Residues: 924-1264,1271-1691 <ZHZ>  
A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459;  
8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:q177922; PIDN:AAA51538.1;  
R:Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.  
Kidney Int. 44, 1316-1321, 1993  
A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a comp  
A:Reference number: I56971; MUID:94133540  
A:Accession: I56971  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1258-1276 <GU03>  
A:Cross-references: GB:S69168; NID:q545095; PIDN:AA060612.1; PID:q545096  
A:Note: kidney splice form  
A:Accession: I76598  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1284-1291,'TFLGVLACLV' <GU02>  
A:Cross-references: GB:S69169; NID:q545097; PIDN:AA060613.1; PID:q545098  
A:Note: frameshift mutation in patient with Alport syndrome  
R:Myers, J.C.; Jones, T.A.; Pohjolaenen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.;  
Am. J. Hum. Genet. 46, 1024-1033, 1990  
A:Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the r  
A:Reference number: A35335; MUID:90252791  
A:Accession: A35335  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1448-1477 <MYE>  
R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Kotabash, Y.; Takada, T.;  
Kidney Int. 46, 1307-1314, 1994  
A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in pr  
A:Reference number: I56975; MUID:95156893  
A:Accession: I56975  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1595-1602 <NAX>  
A:Cross-references: GB:S75903; NID:q913882; PIDN:AA033374.1; PID:q913883  
A:Note: premature termination mutation from a patient with Alport syndrome; one other  
R:Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K  
Genomics 17, 485-489, 1993  
A:Title: Identification of four novel mutations in the COL4A5 gene of patients with A  
A:Reference number: I54188; MUID:94010948  
A:Accession: I54188  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1604-1607,'VHDAYKC' <LEM>  
A:Cross-references: GB:S65767; NID:q425563; PIDN:AA013967.1; PID:q4261667  
A:Note: frameshift mutation from a patient with Alport syndrome; five other mutations  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
ed and subsequently O-glycosylated.  
C:Genetics: GDB:COL4A5; AFS

A:Cross-references: GDB:120596; OMIM:303630  
A:Map position: Xq22-Xq22  
A:Introns: 27/3: 472/3: 77/3: 92/3: 107/3: 128/3: 146/3: 155/3: 182/3: 203/3: 215/3: 229/3: 799/1: 837/1: 893/1: 923/1: 973/1: 1006/1: 1036/1: 1082/3: 1125/1: 1152/1: 1185/1: 1216/1: 1257/1: 1298/1: 1339/1: 1380/1: 1421/1: 1462/1: 1503/1: 1544/1: 1585/1: 1626/1: 1667/1: 1708/1: 1749/1: 1790/1: 1831/1: 1872/1: 1913/1: 1954/1: 1995/1: 2036/1: 2077/1: 2118/1: 2159/1: 2200/1: 2241/1: 2282/1: 2323/1: 2364/1: 2405/1: 2446/1: 2487/1: 2528/1: 2569/1: 2610/1: 2651/1: 2692/1: 2733/1: 2774/1: 2815/1: 2856/1: 2897/1: 2938/1: 2979/1: 3020/1: 3061/1: 3102/1: 3143/1: 3184/1: 3225/1: 3266/1: 3307/1: 3348/1: 3389/1: 3430/1: 3471/1: 3512/1: 3553/1: 3594/1: 3635/1: 3676/1: 3717/1: 3758/1: 3799/1: 3840/1: 3881/1: 3922/1: 3963/1: 4004/1: 4045/1: 4086/1: 4127/1: 4168/1: 4209/1: 4250/1: 4291/1: 4332/1: 4373/1: 4414/1: 4455/1: 4496/1: 4537/1: 4578/1: 4619/1: 4660/1: 4701/1: 4742/1: 4783/1: 4824/1: 4865/1: 4906/1: 4947/1: 4988/1: 5029/1: 5070/1: 5111/1: 5152/1: 5193/1: 5234/1: 5275/1: 5316/1: 5357/1: 5398/1: 5439/1: 5480/1: 5521/1: 5562/1: 5603/1: 5644/1: 5685/1: 5726/1: 5767/1: 5808/1: 5849/1: 5890/1: 5931/1: 5972/1: 6013/1: 6054/1: 6095/1: 6136/1: 6177/1: 6218/1: 6259/1: 6300/1: 6341/1: 6382/1: 6423/1: 6464/1: 6505/1: 6546/1: 6587/1: 6628/1: 6669/1: 6710/1: 6751/1: 6792/1: 6833/1: 6874/1: 6915/1: 6956/1: 6997/1: 7038/1: 7079/1: 7120/1: 7161/1: 7202/1: 7243/1: 7284/1: 7325/1: 7366/1: 7407/1: 7448/1: 7489/1: 7530/1: 7571/1: 7612/1: 7653/1: 7694/1: 7735/1: 7776/1: 7817/1: 7858/1: 7899/1: 7940/1: 7981/1: 8022/1: 8063/1: 8104/1: 8145/1: 8186/1: 8227/1: 8268/1: 8309/1: 8350/1: 8391/1: 8432/1: 8473/1: 8514/1: 8555/1: 8596/1: 8637/1: 8678/1: 8719/1: 8760/1: 8801/1: 8842/1: 8883/1: 8924/1: 8965/1: 9006/1: 9047/1: 9088/1: 9129/1: 9170/1: 9211/1: 9252/1: 9293/1: 9334/1: 9375/1: 9416/1: 9457/1: 9498/1: 9539/1: 9580/1: 9621/1: 9662/1: 9703/1: 9744/1: 9785/1: 9826/1: 9867/1: 9908/1: 9949/1: 9990/1: 10031/1: 10072/1: 10113/1: 10154/1: 10195/1: 10236/1: 10277/1: 10318/1: 10359/1: 10400/1: 10441/1: 10482/1: 10523/1: 10564/1: 10605/1: 10646/1: 10687/1: 10728/1: 10769/1: 10810/1: 10851/1: 10892/1: 10933/1: 10974/1: 11015/1: 11056/1: 11097/1: 11138/1: 11179/1: 11220/1: 11261/1: 11302/1: 11343/1: 11384/1: 11425/1: 11466/1: 11507/1: 11548/1: 11589/1: 11630/1: 11671/1: 11712/1: 11753/1: 11794/1: 11835/1: 11876/1: 11917/1: 11958/1: 12000/1: 12041/1: 12082/1: 12123/1: 12164/1: 12205/1: 12246/1: 12287/1: 12328/1: 12369/1: 12410/1: 12451/1: 12492/1: 12533/1: 12574/1: 12615/1: 12656/1: 12697/1: 12738/1: 12779/1: 12820/1: 12861/1: 12902/1: 12943/1: 12984/1: 13025/1: 13066/1: 13107/1: 13148/1: 13189/1: 13230/1: 13271/1: 13312/1: 13353/1: 13394/1: 13435/1: 13476/1: 13517/1: 13558/1: 13599/1: 13640/1: 13681/1: 13722/1: 13763/1: 13804/1: 13845/1: 13886/1: 13927/1: 13968/1: 14009/1: 14050/1: 14091/1: 14132/1: 14173/1: 14214/1: 14255/1: 14296/1: 14337/1: 14378/1: 14419/1: 14460/1: 14501/1: 14542/1: 14583/1: 14624/1: 14665/1: 14706/1: 14747/1: 14788/1: 14829/1: 14870/1: 14911/1: 14952/1: 14993/1: 15034/1: 15075/1: 15116/1: 15157/1: 15198/1: 15239/1: 15280/1: 15321/1: 15362/1: 15403/1: 15444/1: 15485/1: 15526/1: 15567/1: 15608/1: 15649/1: 15690/1: 15731/1: 15772/1: 15813/1: 15854/1: 15895/1: 15936/1: 15977/1: 16018/1: 16059/1: 16100/1: 16141/1: 16182/1: 16223/1: 16264/1: 16305/1: 16346/1: 16387/1: 16428/1: 16469/1: 16510/1: 16551/1: 16592/1: 16633/1: 16674/1: 16715/1: 16756/1: 16797/1: 16838/1: 16879/1: 16920/1: 16961/1: 17002/1: 17043/1: 17084/1: 17125/1: 17166/1: 17207/1: 17248/1: 17289/1: 17330/1: 17371/1: 17412/1: 17453/1: 17494/1: 17535/1: 17576/1: 17617/1: 17658/1: 17699/1: 17740/1: 17781/1: 17822/1: 17863/1: 17904/1: 17945/1: 17986/1: 18027/1: 18068/1: 18109/1: 18150/1: 18191/1: 18232/1: 18273/1: 18314/1: 18355/1: 18396/1: 18437/1: 18478/1: 18519/1: 18560/1: 18601/1: 18642/1: 18683/1: 18724/1: 18765/1: 18806/1: 18847/1: 18888/1: 18929/1: 18970/1: 19011/1: 19052/1: 19093/1: 19134/1: 19175/1: 19216/1: 19257/1: 19298/1: 19339/1: 19380/1: 19421/1: 19462/1: 19503/1: 19544/1: 19585/1: 19626/1: 19667/1: 19708/1: 19749/1: 19790/1: 19831/1: 19872/1: 19913/1: 19954/1: 19995/1: 20036/1: 20077/1: 20118/1: 20159/1: 20200/1: 20241/1: 20282/1: 20323/1: 20364/1: 20405/1: 20446/1: 20487/1: 20528/1: 20569/1: 20610/1: 20651/1: 20692/1: 20733/1: 20774/1: 20815/1: 20856/1: 20897/1: 20938/1: 20979/1: 21020/1: 21061/1: 21102/1: 21143/1: 21184/1: 21225/1: 21266/1: 21307/1: 21348/1: 21389/1: 21430/1: 21471/1: 21512/1: 21553/1: 21594/1: 21635/1: 21676/1: 21717/1: 21758/1: 21799/1: 21840/1: 21881/1: 21922/1: 21963/1: 21999/1: 22040/1: 22081/1: 22122/1: 22163/1: 22204/1: 22245/1: 22286/1: 22327/1: 22368/1: 22409/1: 22450/1: 22491/1: 22532/1: 22573/1: 22614/1: 22655/1: 22696/1: 22737/1: 22778/1: 22819/1: 22860/1: 22901/1: 22942/1: 22983/1: 23024/1: 23065/1: 23106/1: 23147/1: 23188/1: 23229/1: 23270/1: 23311/1: 23352/1: 23393/1: 23434/1: 23475/1: 23516/1: 23557/1: 23598/1: 23639/1: 23680/1: 23721/1: 23762/1: 23803/1: 23844/1: 23885/1: 23926/1: 23967/1: 24008/1: 24049/1: 24090/1: 24131/1: 24172/1: 24213/1: 24254/1: 24295/1: 24336/1: 24377/1: 24418/1: 24459/1: 24500/1: 24541/1: 24582/1: 24623/1: 24664/1: 24705/1: 24746/1: 24787/1: 24828/1: 24869/1: 24910/1: 24951/1: 24992/1: 25033/1: 25074/1: 25115/1: 25156/1: 25197/1: 25238/1: 25279/1: 25320/1: 25361/1: 25402/1: 25443/1: 25484/1: 25525/1: 25566/1: 25607/1: 25648/1: 25689/1: 25730/1: 25771/1: 25812/1: 25853/1: 25894/1: 25935/1: 25976/1: 26017/1: 26058/1: 26099/1: 26140/1: 26181/1: 26222/1: 26263/1: 26304/1: 26345/1: 26386/1: 26427/1: 26468/1: 26509/1: 26550/1: 26591/1: 26632/1: 26673/1: 26714/1: 26755/1: 26796/1: 26837/1: 26878/1: 26919/1: 26960/1: 27001/1: 27042/1: 27083/1: 27124/1: 27165/1: 27206/1: 27247/1: 27288/1: 27329/1: 27370/1: 27411/1: 27452/1: 27493/1: 27534/1: 27575/1: 27616/1: 27657/1: 27698/1: 27739/1: 27780/1: 27821/1: 27862/1: 27903/1: 27944/1: 27985/1: 28026/1: 28067/1: 28108/1: 28149/1: 28190/1: 28231/1: 28272/1: 28313/1: 28354/1: 28395/1: 28436/1: 28477/1: 28518/1: 28559/1: 28600/1: 28641/1: 28682/1: 28723/1: 28764/1: 28805/1: 28846/1: 28887/1: 28928/1: 28969/1: 29010/1: 29051/1: 29092/1: 29133/1: 29174/1: 29215/1: 29256/1: 29297/1: 29338/1: 29379/1: 29420/1: 29461/1: 29502/1: 29543/1: 29584/1: 29625/1: 29666/1: 29707/1: 29748/1: 29789/1: 29830/1: 29871/1: 29912/1: 29953/1: 29994/1: 30035/1: 30076/1: 30117/1: 30158/1: 30199/1: 30240/1: 30281/1: 30322/1: 30363/1: 30404/1: 30445/1: 30486/1: 30527/1: 30568/1: 30609/1: 30650/1: 30691/1: 30732/1: 30773/1: 30814/1: 30855/1: 30896/1: 30937/1: 30978/1: 31019/1: 31060/1: 31101/1: 31142/1: 31183/1: 31224/1: 31265/1: 31306/1: 31347/1: 31388/1: 31429/1: 31470/1: 31511/1: 31552/1: 31593/1: 31634/1: 31675/1: 31716/1: 31757/1: 31798/1: 31839/1: 31880/1: 31921/1: 31962/1: 32003/1: 32044/1: 32085/1: 32126/1: 32167/1: 32208/1: 32249/1: 32290/1: 32331/1: 32372/1: 32413/1: 32454/1: 32495/1: 32536/1: 32577/1: 32618/1: 32659/1: 32700/1: 32741/1: 32782/1: 32823/1: 32864/1: 32905/1: 32946/1: 32987/1: 33028/1: 33069/1: 33110/1: 33151/1: 33192/1: 33233/1: 33274/1: 33315/1: 33356/1: 33397/1: 33438/1: 33479/1: 33520/1: 33561/1: 33602/1: 33643/1: 33684/1: 33725/1: 33766/1: 33807/1: 33848/1: 33889/1: 33930/1: 33971/1: 34012/1: 34053/1: 34094/1: 34135/1: 34176/1: 34217/1: 34258/1: 34299/1: 34340/1: 34381/1: 34422/1: 34463/1: 34504/1: 34545/1: 34586/1: 34627/1: 34668/1: 34709/1: 34750/1: 34791/1: 34832/1: 34873/1: 34914/1: 34955/1: 34996/1: 35037/1: 35078/1: 35119/1: 35160/1: 35201/1: 35242/1: 35283/1: 35324/1: 35365/1: 35406/1: 35447/1: 35488/1: 35529/1: 35570/1: 35611/1: 35652/1: 35693/1: 35734/1: 35775/1: 35816/1: 35857/1: 35898/1: 35939/1: 35980/1: 36021/1: 36062/1: 36103/1: 36144/1: 36185/1: 36226/1: 36267/1: 36308/1: 36349/1: 36390/1: 36431/1: 36472/1: 36513/1: 36554/1: 36595/1: 36636/1: 36677/1: 36718/1: 36759/1: 36800/1: 36841/1: 36882/1: 36923/1: 36964/1: 37005/1: 37046/1: 37087/1: 37128/1: 37169/1: 37210/1: 37251/1: 37292/1: 37333/1: 37374/1: 37415/1: 37456/1: 37497/1: 37538/1: 37579/1: 37620/1: 37661/1: 37702/1: 37743/1: 37784/1: 37825/1: 37866/1: 37907/1: 37948/1: 37989/1: 38030/1: 38071/1: 38112/1: 38153/1: 38194/1: 38235/1: 38276/1: 38317/1: 38358/1: 38399/1: 38440/1: 38481/1: 38522/1: 38563/1: 38604/1: 38645/1: 38686/1: 38727/1: 38768/1: 38809/1: 38850/1: 38891/1: 38932/1: 38973/1: 39014/1: 39055/1: 39096/1: 39137/1: 39178/1: 39219/1: 39260/1: 39301/1: 39342/1: 39383/1: 39424/1: 39465/1: 39506/1: 39547/1: 39588/1: 39629/1: 39670/1: 39711/1: 39752/1: 39793/1: 39834/1: 39875/1: 39916/1: 39957/1: 40000/1: 40041/1: 40082/1: 40123/1: 40164/1: 40205/1: 40246/1: 40287/1: 40328/1: 40369/1: 40410/1: 40451/1: 40492/1: 40533/1: 40574/1: 40615/1: 40656/1: 40697/1: 40738/1: 40779/1: 40820/1: 40861/1: 40902/1: 40943/1: 40984/1: 41025/1: 41066/1: 41107/1: 41148/1: 41189/1: 41230/1: 41271/1: 41312/1: 41353/1: 41394/1: 41435/1: 41476/1: 41517/1: 41558/1: 41599/1: 41640/1: 41681/1: 41722/1: 41763/1: 41804/1: 41845/1: 41886/1: 41927/1: 41968/1: 42009/1: 42050/1: 42091/1: 42132/1: 42173/1: 42214/1: 42255/1: 42296/1: 42337/1: 42378/1: 42419/1: 42460/1: 42501/1: 42542/1: 42583/1: 42624/1: 42665/1: 42706/1: 42747/1: 42788/1: 42829/1: 42870/1: 42911/1: 42952/1: 42993/1: 43034/1: 43075/1: 43116/1: 43157/1: 43198/1: 43239/1: 43280/1: 43321/1: 43362/1: 43403/1: 43444/1: 43485/1: 43526/1: 43567/1: 43608/1: 43649/1: 43690/1: 43731/1: 43772/1: 43813/1: 43854/1: 43895/1: 43936/1: 43977/1: 44018/1: 44059/1: 44100/1: 44141/1: 44182/1: 44223/1: 44264/1: 44305/1: 44346/1: 44387/1: 44428/1: 44469/1: 44510/1: 44551/1: 44592/1: 44633/1: 44674/1: 44715/1: 44756/1: 44797/1: 44838/1: 44879/1: 44920/1: 44961/1: 45002/1: 45043/1: 45084/1: 45125/1: 45166/1: 45207/1: 45248/1: 45289/1: 45330/1: 45371/1: 45412/1: 45453/1: 45494/1: 45535/1: 45576/1: 45617/1: 45658/1: 45699/1: 45740/1: 45781/1: 45822/1: 45863/1: 45904/1: 45945/1: 45986/1: 46027/1: 46068/1: 46109/1: 46150/1: 46191/1: 46232/1: 46273/1: 46314/1: 46355/1: 46396/1: 46437/1: 46478/1: 46519/1: 46560/1: 46601/1: 46642/1: 46683/1: 46724/1: 46765/1: 46806/1: 46847/1: 46888/1: 46929/1: 46970/1: 47011/1: 47052/1: 47093/1: 47134/1: 47175/1: 47216/1: 47257/1: 47298/1: 47339/1: 47380/1: 47421/1: 47462/1: 47503/1: 47544/1: 47585/1: 47626/1: 47667/1: 47708/1: 47749/1: 47790/1: 47831/1: 47872/1: 47913/1: 47954/1: 47995/1: 48036/1: 48077/1: 48118/1: 48159/1: 48200/1: 48241/1: 48282/1: 48323/1: 48364/1: 48405/1: 48446/1: 48487/1: 48528/1: 48569/1: 48610/1: 48651/1: 48692/1: 48733/1: 48774/1: 48815/1: 48856/1: 48897/1: 48938/1: 48979/1: 49020/1: 49061/1: 49102/1: 49143/1: 49184/1: 49225/1: 49266/1: 49307/1: 49348/1: 49389/1: 49430/1: 49471/1: 49512/1: 49553/1: 49594/1: 49635/1: 49676/1: 49717/1: 49758/1: 49799/1: 49840/1: 49881/1: 49922/1: 49963/1: 50004/1: 50045/1: 50086/1: 50127/1: 50168/1: 50209/1: 50250/1: 50291/1: 50332/1: 50373/1: 50414/1: 50455/1: 50496/1: 50537/1: 50578/1: 50619/1: 50660/1: 50701/1: 50742/1: 50783/1: 50824/1: 50865/1: 50906/1: 50947/1: 50988/1: 51029/1: 51070/1: 51111/1: 51152/1: 51193/1: 51234/1: 51275/1: 51316/1: 51357/1: 51398/1: 51439/1: 51480/1: 51521/1: 51562/1: 51603/1: 51644/1: 51685/1: 51726/1: 51767/1: 51808/1: 51849/1: 51890/1: 51931/1: 51972/1: 51999/1: 52026/1: 52053/1: 52080/1: 52107/1: 52134/1: 52161/1: 52188/1: 52215/1: 52242/1: 52269/1: 52296/1: 52323/1: 52350/1: 52377/1: 52404/1: 52431/1: 52458/1: 52485/1: 52512/1: 52539/1: 52566/1: 52593/1: 52620/1: 52647/1: 52674/1: 52701/1: 52728/1: 52755/1: 52782/1: 52809/1: 52836/1: 52863/1: 52890/1: 52917/1: 52944/1: 52971/1: 53000/1: 53027/1: 53054/1: 53081/1: 53108/1: 53135/1: 53162/1: 53189/1: 53216/1: 53243/1: 53270/1: 53297/1: 53324/1: 53351/1: 53378/1: 53405/1: 53432/1: 53459/1: 53486/1: 53513/1: 53540/1: 53567/1: 53594/1: 53621/1: 53648/1: 53675/1: 53702/1: 53729/1: 53756/1: 53783/1: 53810/1: 53837/1: 53864/1: 53891/1: 53918/1: 53945/1: 53972/1: 54000/1: 54027/1: 54054/1: 54081/1: 54108/1: 54135/1: 54162/1: 54189/1: 54216/1: 54243/1: 54270/1: 54297/1: 54324/1: 54351/1: 54378/1: 54405/1: 54432/1: 54459/1: 54486/1: 54513/1: 54540/1: 54567/1: 54594/1: 54621/1: 54648/1: 54675/1: 54702/1: 54729/1: 54756/1: 54783/1: 54810/1: 54837/1: 54864/1: 54891/1: 54918/1: 54945/1: 54972/1: 55000/1: 55027/1: 55054/1: 55081/1: 55108/1: 55135/1: 55162/1: 55189/1: 55216/1:

A:Residues: 1-60 <BR>  
A:Cross-references: EMBL:M23333  
R:Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 139, 401-410, 1984  
A:Title: Subunit structure and assembly of the globular domain of basement-membrane coll  
A:Reference number: S17801; MUID:84132058  
A:Accession: S19086  
A:Molecule type: protein  
A:Residues: 1475-1481, 'X', 1483-1487 <WEB>  
C:Genetics:  
A:Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>  
F:29-171/Domain: 7S #status predicted <7SD>  
F:58-1480/Domain: collagenous #status predicted <COL>  
F:141-143/Region: cell attachment (R-G-D) motif  
F:360-362/Region: cell attachment (R-G-D) motif  
F:779-781/Region: cell attachment (R-G-D) motif  
F:884-886/Region: cell attachment (R-G-D) motif  
F:965-967/Region: cell attachment (R-G-D) motif  
F:1223-1225/Region: cell attachment (R-G-D) motif  
F:1447-1449/Region: cell attachment (R-G-D) motif  
F:1481-1707/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F:1590-1707/Domain: repeat NC1 #status predicted <NC12>  
F:42,47,51,53,481,483/Disulfide bonds: interchain #status predicted  
F:138,1270/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:656,676,1544-1550,1653-1660/Disulfide bonds: #status predicted  
F:988,988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Bind  
F:988,988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Modi  
F:1009,1012,1018,1021,1024,1037,1040,1043,1046,1052,1058,1070,1098,1110,1128,1140,1149,1  
77,1383,1386,1401,1408,1420,1423,1429,1444,1465,1466,1471,1477/Modified site: hydroxypro

Query Match 70.7%; Score 65; DB 2; Length 1707;  
Best Local Similarity 66.7%; Pred. No. 0.16;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNPWPAP 15  
1:|||||  
Db 1165 GIPGDKDPGPGVP 1179

RESULT 7  
CGHUB  
N:Alternate names: procollagen alpha 2(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1990 #sequence, revision 03-Oct-1995 #text-change 22-Jun-1999  
C:Accession: A32024; S00007; S02624; S00246; S17678; S16911; B32117; S16877; S00165; S39  
R:Hostlikka, S.L.; Tryggvason, K.  
J. Biol. Chem. 263, 19488-19493, 1988  
A:Title: The complete primary structure of the alpha2 chain of human type IV collagen an  
A:Reference number: A32024; MUID:89066769  
A:Accession: A32024  
A:Molecule type: mRNA  
A:Residues: 1-1712 <HOS1>  
A:Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:929550; PIDN:CAA29098.1; PI  
R:Hostlikka, S.L.; Kukkien, M.; Tryggvason, K.  
FEBS Lett. 216, 281-286, 1987  
A:Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA re  
ated region.  
A:Reference number: S00007; MUID:87219158  
A:Accession: S00007  
A:Molecule type: mRNA  
A:Residues: 1254-1398, 'V', 1400-1712 <HOS2>  
A:Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:929550; PIDN:CAA29098.1; PI  
A:Note: 1399-ile was also found  
R:Hostlikka, S.L.; Tryggvason, K.  
FEBS Lett. 224, 297-305, 1987  
A:Title: Extensive structural differences between genes for the alpha(1) and alpha(2) ch  
A:Reference number: S02624; MUID:88083553

A:Accession: S02624  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1347-1350,1377-1383,1426-1432,1465-1471,1529-1535,1625-1630 <HOS3>  
A:Note: complete nucleotide sequence not shown  
R:Brizel, D.; Pollner, R.; Oberhaemer, I.; Kuehn, K.  
Eur. J. Biochem. 172, 35-42, 1988  
A:Title: Human basement membrane collagen (type IV): the amino acid sequence of the a  
A:Reference number: S00246; MUID:88151998  
A:Accession: S00246  
A:Molecule type: mRNA  
A:Residues: 1-682, 'G', 684-1043 <BR>  
A:Cross-references: EMBL:X05562; NID:930075; PIDN:CAA29076.1; PID:930076  
R:Oberhaemer, I.  
submitted to the EMBL Data Library, June 1987  
A:Reference number: S17678  
A:Accession: S17678  
A:Molecule type: mRNA  
A:Residues: 1-470, 'P', 472-682, 'G', 684-1043 <OB>  
A:Cross-references: EMBL:X05562; NID:930075; PIDN:CAA29076.1; PID:930076  
R:Poeschl, E.; Pollner, R.; Kuehn, K.  
EMBO J. 7, 2687-2695, 1988  
A:Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membran  
A:Reference number: S02738; MUID:89030632  
A:Accession: S16911  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 1-33 <POE>  
A:Cross-references: EMBL:X12784; GB:M36963; NID:930072; PIDN:CAA31275.1; PID:930073  
R:Solninen, R.; Huotari, M.; Hostlikka, S.L.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 263, 17217-17220, 1988  
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen  
A:Reference number: A92690; MUID:89034231  
A:Accession: B32117  
A:Molecule type: DNA  
A:Residues: 1-33 <SO11>  
A:Cross-references: EMBL:J04217; EMBL:J05039; NID:918075; PIDN:AA53097.1; PID:95532  
R:Solninen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.  
J. Biol. Chem. 264, 13565-13571, 1989  
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV c  
A:Reference number: S16876; MUID:89340433  
A:Accession: S16877  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-33 <SO12>  
A:Cross-references: EMBL:J04217; NID:918075; PIDN:AA53097.1; PID:95532  
R:Stebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, H.; Deutmann, R.; Kuehn, K.  
Eur. J. Biochem. 168, 569-575, 1987  
A:Title: Construction of a model for the aggregation and cross-linking region (7S dom  
is region.  
A:Reference number: S00165; MUID:88029476  
A:Accession: S00165  
A:Molecule type: protein  
A:Residues: 37-247 <STE1>  
A:Note: The sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-  
EMBO J. 12, 4795-4802, 1993  
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collag  
A:Reference number: S39614; MUID:94038963  
A:Accession: S39615  
A:Molecule type: protein  
A:Residues: 407-570 <RBL>  
R:MacRlight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.  
Biochemistry 22, 4940-4948, 1983  
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane  
A:Reference number: S16910; MUID:84053346  
A:Accession: S16912  
A:Molecule type: protein  
A:Residues: 490-492, 'X', 494-496, 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>  
A:Experimental source: placenta  
R:Glanville, R.W.; Rauter, A.  
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981





A:Accession: A28518  
A:Molecule type: protein  
A:Residues: 156-159, 'X', 161-166, 'X', 168-173, 'X', 175-178 <HUB>  
J. Rivan der Rest, M.; Mayne, R.  
J. Biol. Chem. 263, 1615-1618, 1988  
A:Title: Type IX collagen proteoglycan from cartilage is covalently cross-linked to type  
A:Reference number: S23813; MUID:88115274  
A:Accession: S23813  
A:Molecule type: protein  
A:Residues: 170-180, 'X', 182-184 <VAN>  
A:Note: evidence for aldimine cross-linkage of 190-Lys to collagen alpha 1(VII) chain is  
R. Mayne, R.; van der Rest, M.; Nimmiya, Y.; Olsen, B.R.  
Ann. N. Y. Acad. Sci. 460, 38-46, 1985  
A:Title: The structure of type IX collagen  
A:Reference number: S22238; MUID:86185164  
A:Accession: S22239  
A:Molecule type: protein  
A:Residues: 542-567 <MAT>  
C:Genetics:  
A:Introns: 427/3; 439/3; 454/3; 465/3; 514/3; 533/1; 596/1; 622/1  
C:Superfamily: unassigned collagens  
C:Keywords: chondroitin sulfate proteoglycan; coll; extracellular matrix; glycop  
F:1-21/Domains: signal sequence #status predicted <SIG>  
F:22-677/Product: collagen alpha 2(XI) chain #status predicted <MAT>  
F:23-161/Domains: collagenous COL3 #status predicted <COL3>  
F:162-178/Domains: non-collagenous NC3 #status predicted <NC3>  
F:179-517/Domains: collagenous COL2 #status predicted <COL2>  
F:518-547/Domains: non-collagenous NC2 #status predicted <NC2>  
F:548-662/Domains: collagenous COL1 #status predicted <COL1>  
F:663-677/Domains: non-collagenous NC1 #status predicted <NC1>  
F:138,178/Modified site: 4-hydroxyproline (Pro) #status experimental  
F:167/Binding site: Chondroitin sulfate (Ser) (covalent) #status experimental  
F:181/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:181/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F:190/Modified site: allysine (Lys) #status predicted

Query Match 68.5%; Score 63; DB 2; Length 677;  
Best Local Similarity 66.7%; Pred. No. 0.13;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GVKDGKNGMPCAP 15  
Db 46 GIDGKSGFAGPSP 60

RESULT 9  
A54849  
collagen alpha 1(VII) chain precursor - human  
N:Alternate names: procollagen alpha 1(VII) chain  
C:Species: Homo sapiens (man)  
C:Date: 04-Nov-1994 #sequence\_revision 04-Nov-1994 #text\_change 20-Sep-1999  
R:Accession: A54849; S16316; S16328; A30296; I84686  
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ultio, J.  
J. Biol. Chem. 269, 20256-20262, 1994  
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)  
A:Reference number: A54849; MUID:94327588  
A:Accession: A54849  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-2944 <CHR>  
A:Cross-references: GB:L02870; NID:9987124; PIDN:AAA75438.1; PID:9987125  
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A:Title: Molecular cloning and characterization of type VII collagen cDNA.  
A:Reference number: PH0844; MUID:92231902  
A:Accession: PH0844  
A:Molecule type: mRNA  
A:Residues: 'ERR', 340-475, 'BALSTASHSTICWATRMWPCNRGSHMTACAPCPNRPASRARAG', 524-528, 'C',  
A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BA02853.1; PID:9453699  
A:Experimental source: keratinocyte  
A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
R:Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A:Reference number: S16316; MUID:91334380  
A:Accession: S16316  
A:Molecule type: mRNA  
A:Residues: 815-892, 'E', 894-1439 <PAR>  
A:Cross-references: GB:M65156; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
A:Experimental source: keratinocyte  
R:Gammou, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisyayanh, P.S.; Cook, M.E.; Wright,  
J. Invest. Dermatol. 99, 691-696, 1992  
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion p  
A:Reference number: I56328; MUID:93107742  
A:Accession: I56328  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 'ERR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
A:Cross-references: GB:S51236; NID:g262308; PIDN:AA824637.1; PID:g262309  
R:Seitzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Gnanville, R.W.; Burgeson, R  
J. Biol. Chem. 264, 3822-3826, 1989  
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV collag  
A:Reference number: A30296; MUID:89139437  
A:Accession: A30296  
A:Molecule type: protein  
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-204  
A:Note: two reported peptides cannot be reliably located  
R:Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collageno  
A:Reference number: I48103; MUID:93271985  
A:Accession: I48103  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 2395-2871, 'S', 2873-2944 <RE2>  
A:Cross-references: GB:L06862; NID:g388713; PIDN:AAA89196.1; PID:g388714  
R:Christiano, A.M.; Ryyanen, M.; Ultio, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser s  
A:Reference number: A5255; MUID:94224777  
A:Accession: A5255  
A:Contents: annotation  
A:Comment: Prolines and lysines at the third position of the tripeptide repeating uni  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL7A1; EBR1; EBD1; EB  
A:Cross-references: GDB:128750; OMIM:120120  
A:Map position: 3p21.3-3p21.3  
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermo  
A:Note: there are 118 introns  
C:Complex: type VII collagen is probably a homotrimer  
C:Function:  
A:Description: structural component of extracellular polymer associated with anchorin  
C:Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homology  
C:Keywords: coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxypr  
F:1-16/Domains: signal sequence #status predicted <SIG>  
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F:17-1255/Domains: amino-terminal nonhelical #status predicted <NC1>  
F:36-201/Domains: von Willebrand factor type A repeat homology <VWA1>  
F:231-318/Domains: fibronectin type III repeat homology <FN1>  
F:327-413/Domains: fibronectin type III repeat homology <FN2>  
F:414-502/Domains: fibronectin type III repeat homology <FN3>  
F:508-593/Domains: fibronectin type III repeat homology <FN4>  
F:598-683/Domains: fibronectin type III repeat homology <FN5>  
F:688-771/Domains: fibronectin type III repeat homology <FN6>  
F:776-862/Domains: fibronectin type III repeat homology <FN7>  
F:864-952/Domains: fibronectin type III repeat homology <FN8>  
F:954-1045/Domains: fibronectin type III repeat homology <FN9>  
F:1052-1219/Domains: von Willebrand factor type A repeat homology <VWA2>  
F:1170-1172/Domains: cell attachment (R-G-D) motif  
F:1189-1253/Domains: cysteine/proline-rich  
F:1254-2783/Domains: interrupted helical  
F:1334-1336/Domains: cell attachment (R-G-D) motif  
F:2008-2010/Domains: cell attachment (R-G-D) motif  
F:2553-2555/Domains: cell attachment (R-G-D) motif  
F:2784-2944/Domains: carboxyl-terminal nonhelical #status predicted <NC2>  
F:2876-2929/Domains: animal Kunitz-type proteinase inhibitor homology <BPI>



F:337,786,1109/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:2167,2116,2188,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status expected  
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:2625,2631/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F:2634,2802,2804/Disulfide bonds: Interchain #status predicted

Query Match  
Best Local Similarity 68.5%; Score 63; DB 2; Length 2944;  
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGMPCGAP 15  
Db 2386 GVKGDLGLPCLPGAP 2400

RESULT 10  
A32249  
collagen - sea urchin (Paracentrotus lividus) (fragment)  
C:Species: Paracentrotus lividus (common urchin)  
C>Date: 17-Aug-1989 #sequence\_revision 17-Aug-1989 #text\_change 19-Jan-1996  
C:Accession: A32249

R:Saltta, B.; Buttice, G.; Gambino, R.  
Blochm, Biophys. Res. Commun. 158, 633-639, 1989  
A:Title: Isolation of a putative collagen-like gene from the sea urchin Paracentrotus lividus  
A:Reference number: A32249; MWID:89149773  
A:Accession: A32249

A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-290 <SAIS>  
C:Superfamily: 1-290 <SAIS>  
C:Keywords: collid coll; extracellular matrix; glycoprotein; trimer; triple helix

Query Match  
Best Local Similarity 67.4%; Score 62; DB 2; Length 290;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGMPCGAP 14  
Db 76 GSKGDCGNPGMPCGAP 89

RESULT 11  
S23809  
collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)  
C:Species: Strongylocentrotus purpuratus (purple urchin)  
C>Date: 10-Sep-1989 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: S23809

R:Exposito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.  
J. Biol. Chem. 267, 15559-15562, 1992  
A:Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) collagen  
A:Reference number: S23809; MWID:92338411  
A:Accession: S23809

A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1414 <EXP>  
C:Cross-references: EMBL:M92040; NID:g161435; PIDN:AAA30035.1; PID:g161436  
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology  
C:Keywords: collid coll; extracellular matrix; glycoprotein; trimer; triple helix  
F:1207-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCG>

Query Match  
Best Local Similarity 67.4%; Score 62; DB 1; Length 1414;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGMPCGAP 14  
Db 393 GSKGDCGNPGMPCGAP 406

RESULT 12  
CGH038  
collagen alpha 3(IV) chain precursor, long splice form - human  
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain; long splice form

C:Species: Homo sapiens (man)  
C>Date: 28-Oct-1994 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C:Accession: A54763; A43928; A44043; A45971; A39786  
R:Marlyama, M.; Leinonen, A.; Mochizuki, T.; Tytgvaason, K.; Reeders, S.T.  
J. Biol. Chem. 269, 23013-23017, 1994  
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress  
A:Reference number: A54763; MWID:94364994  
A:Accession: A54763

A:Molecule type: mRNA  
A:Residues: 1-1670 <MAN>  
A:Cross-references: GB:X80031; NID:9577563; PID:9577564  
A:Experimental source: Kidney  
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.  
J. Clin. Invest. 89, 592-601, 1992  
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the  
A:Reference number: A43928; MWID:92147878  
A:Accession: A43928

A:Molecule type: mRNA  
A:Residues: 1331-1524, 'T', 1526-1670 <TUR>  
A:Cross-references: GB:M81379  
A:Experimental source: Kidney  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 267, 19780-19784, 1992  
A:Title: Exon/Intron structure of the human alpha 3(IV) gene encompassing the Goodpasture antigen.

A:Reference number: A44043; MWID:93015826  
A:Accession: A44043  
A:Molecule type: DNA; mRNA  
A:Residues: 1386-1670 <DUI>  
A:Cross-references: GB:M92993; NID:g177895; PIDN:AAA21610.1; PID:g177896  
A:Note: sequence extracted from NCBI backbone (NCBI:115597)  
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.  
J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; MWID:94274734  
A:Contents: annotation; extratum; correction to intronic sequence in A44043  
R:Bernal, D.; Quinones, S.; Saus, J.  
J. Biol. Chem. 268, 12090-12094, 1993  
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.  
A:Reference number: A45971; MWID:93280184  
A:Accession: A45971

A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1427-1444 <BER>  
A:Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly id  
R:Morrison, K.E.; Marlyama, M.; Yang-Feng, T.L.; Reeders, S.T.  
Am. J. Hum. Genet. 49, 545-554, 1991  
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain  
A:Reference number: A39786; MWID:91353570  
A:Accession: A39786

A:Molecule type: mRNA  
A:Residues: 1453-1593, 'A', 1595-1670 <MOP>  
A:Cross-references: GB:S5790; NID:g234418; PIDN:AA19637.1; PID:g234419  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epit  
C:Genetics:

A:Gene: GDB:COL4A3  
A:Cross-references: GDB:128351; OMIM:120070  
A:Map position: 2q35-2q37  
A:Notes: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete  
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands w  
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha  
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimer  
er associations in the interrupted helical domain (with disulfide and desmosine cross  
C:Function:

A:Description: minor structural component of extracellular basement membrane in kidney  
C:Superfamily: collagen alpha 1(IV) chain  
C:Keywords: alternative splicing; basement membrane; cell binding; collid coll; extra  
F:1-28/Domain: signal sequence #status predicted <SIG>  
F:29-1670/Product: collagen alpha 3(IV) chain; long splice form #status predicted <MAN>  
F:29-42/Domain: amino-terminal nonhelical, NHI <NHI>  
F:43-1438/Region: interrupted helical  
F:791-793/Region: cell attachment (R-G-D) motif

F:996-998/Region: cell attachment (R-G-D) motif  
F:1154-1156/Region: cell attachment (R-G-D) motif  
F:1306-1308/Region: cell attachment (R-G-D) motif  
F:1345-1347/Region: cell attachment (R-G-D) motif  
F:1432-1434/Region: cell attachment (R-G-D) motif  
F:1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F:31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predicted  
F:253/Binding site: carbohydrate (asn) (covalent)#status predicted  
F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted  
F:1505-1511,1616-1622/Disulfide bonds: #status predicted  
F:1570-1662,1604-1663/Disulfide bonds: (or 1570-1663, 1604-1662) #status predicted

Query Match 67.4%; Score 62; DB 1; Length 1670;  
Best Local Similarity 66.7%; Pred. No. 0.44;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPMPGAP 15  
DB 1189 GAKGDRGAPGPGP 1203

RESULT 13  
B46345  
gene V protein - phage PRD1  
N:Alternate names: minor capsid protein V  
C:Species: phage PRD1  
C:Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
C:Accession: B46345; D46345  
R:Bamford, J.K.H.; Bamford, D.H.  
Virolology 177, 445-451, 1990  
A:Title: Capsomer proteins of bacteriophage PRD1, a bacterial virus with a membrane.  
A:Reference number: A46345; MUID:90320115  
A:Accession: B46345  
A:Molecule type: DNA  
A:Residues: 1-340 <BAM>  
A:Cross-references: GB:M5568; NID:g215745; PIDN:AAA32446.1; PID:g215746  
A:Accession: D46345  
A:Molecule type: protein  
A:Residues: 2-6 <BA2>  
C:Genetics:  
A:Gene: V  
C:Superfamily: phage PRD1 gene V protein  
C:Keywords: capsid protein

Query Match 66.3%; Score 61; DB 1; Length 340;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKGNPMPG 13  
DB 124 GIKGDKGDPGAPG 136

RESULT 14  
S23810  
collagen alpha 1(XVI) chain precursor - human  
N:Alternate names: procollagen alpha 1(XVI) chain  
C:Species: Homo sapiens (man)  
C:Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 20-Sep-1999  
C:Accession: S23810; PQ0612; S08012  
R:Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Tampl, R.; Chu, M.L.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992  
A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.  
A:Reference number: S23810; MUID:92335339  
A:Accession: S23810  
A:Molecule type: mRNA  
A:Residues: 1-1603 <PAN>  
A:Cross-references: EMBL:M92642; NID:g180757; PIDN:AAA58427.1; PID:g180758  
A:Experimental source: skin fibroblasts  
R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Horii, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J.; Biochem. 112, 856-863, 1992

A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha 1(XVI)  
A:Reference number: PQ0612; MUID:93203161  
A:Accession: PQ0612  
A:Molecule type: mRNA  
A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>  
A:Cross-references: GB:S57132; NID:g2386641; PIDN:AB25797.1; PID:g2386642  
A:Experimental source: placenta  
R:Kimura, S.  
submitted to the EMBL Data Library, April 1989  
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein  
A:Reference number: S08012  
A:Accession: S08012  
A:Molecule type: mRNA  
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM', <KIM>  
A:Cross-references: EMBL:X14963; NID:g29984; PIDN:CAA33085.1; PID:g9330048  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL16A1  
A:Cross-references: GDB:134045; OMIM:120326  
A:Map position: 1p34-1p34  
A:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XV) C:Function:  
A:Description: structural component of extracellular fibrous polymer as a minor form  
A:Note: may play a role in forming elastic connections at fibril surfaces  
C:Superfamily: unassigned collagens  
C:Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxyls  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAMT>  
F:22-333/Domain: amino-terminal nonhelical #status predicted <NC11>  
F:334-1577/Region: interrupted helical  
F:334-360/Domain: collagenous COL10 #status predicted <COL10>  
F:375-505/Domain: collagenous COL9 #status predicted <COL9>  
F:521-554/Domain: collagenous COL8 #status predicted <COL8>  
F:539-541/Region: cell attachment (R-G-D) motif  
F:572-630/Domain: collagenous COL7 #status predicted <COL7>  
F:655-722/Domain: collagenous COL6 #status predicted <COL6>  
F:738-875/Domain: collagenous COL5 #status predicted <COL5>  
F:887-938/Domain: collagenous COL4 #status predicted <COL4>  
F:973-987/Domain: collagenous COL3 #status predicted <COL3>  
F:1005-1007/Region: cell attachment (R-G-D) motif  
F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>  
F:1226-1228/Region: cell attachment (R-G-D) motif  
F:1472-1577/Domain: collagenous COL1 #status predicted <COL1>  
F:1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NC01>  
F:47,327/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 66.3%; Score 61; DB 2; Length 1603;  
Best Local Similarity 73.3%; Pred. No. 0.59;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GVKGDKGNPMPGAP 15  
DB 661 GOKGDKGNPDPGTP 695

RESULT 15  
CGHUB  
collagen alpha 4(IV) chain precursor - human  
N:Alternate names: procollagen alpha 4(IV) chain  
C:Species: Homo sapiens (man)  
C:Date: 06-Feb-1995 #sequence\_revision 03-Oct-1995 #text\_change 16-Jun-2000  
C:Accession: A55360; S36854; S28777  
R:Leinonen, A.; Mariyama, M.; Mochizuki, T.; Trygvaason, K.; Reiders, S.T.  
J. Biol. Chem. 269, 26172-26177, 1994  
A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. C  
A:Reference number: A55360; MUID:95014445  
A:Accession: A55360  
A:Reference number: A55360  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1690 <LEI>  
A:Cross-references: GB:X01053; NID:g574805; PIDN:CAA56943.1; PID:g574806  
R:Sugimoto, M.; Ohashi, T.; Yoshioka, H.; Matsuo, N.; Nishimura, Y.



Query Match 65.2%; Score 60; DB 2; Length 623;  
Best Local Similarity 76.9%; Pred. No. 0.32;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKDKNPMPGA 13  
|||||:|:| |  
Db 139 GVKGERGSPGPGA 151

RESULT 19

S41067  
collagen alpha 1(III) chain - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text\_change 13-Aug-1999  
C:Accession: S41067; A29905; S31924  
R:Glumoff, V.; Maekela, J.K.; Vuorio, E.  
Biochim. Biophys. Acta 1217, 41-48, 1994  
A:Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pat  
A:Reference number: S41067; MUID:94114571  
A:Accession: S41067  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-636 <GLD>  
A:Cross-references: EMBL:X70369; NID:957915; PIDN:CAA9832.1; PID:957916  
R:Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Iln, E.; Lyttle, C.R.; Komm, B.; Mohn, K.  
DNA 7, 347-354, 1988  
A:Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2(V) collagen mRNAs by est  
A:Reference number: A29905; MUID:88296083  
A:Accession: A29905  
A:Molecule type: mRNA  
A:Residues: 308-482 <FRA>  
A:Cross-references: GB:M1334; NID:9203500; PIDN:AAA40942.1; PID:9203501  
R:Glumoff, V.; Maekela, J.K.; Vuorio, E.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S31924  
A:Accession: S31924  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 2-636 <GL2>  
A:Cross-references: EMBL:X70369  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coll; extracellular matrix; glycoprotein; trimer; triple helix  
F:408-636/domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 65.2%; Score 60; DB 2; Length 636;  
Best Local Similarity 71.4%; Pred. No. 0.33;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDKNPMPGA 14  
|||||:|:| |  
Db 30 GVKGERGSPGPGA 43

RESULT 20

150694  
collagen alpha 1(III) chain - chicken (fragment)  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence revision 13-Sep-1996 #text\_change 13-Aug-1999  
C:Accession: 150694  
R:Nah, H.D.; Niu, Z.; Adams, S.L.  
J. Biol. Chem. 269, 16443-16448, 1994  
A:Title: An alternative transcript of the chick type III collagen gene that does not end  
A:Reference number: A54041; MUID:9426842  
A:Accession: 150694  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-886 <RNAH>  
A:Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:9537432  
C:Genetics:  
A:Gene: COL3A1  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
F:30-90/domain: von Willebrand factor type C repeat homology <VWC>

Query Match 65.2%; Score 60; DB 2; Length 886;  
Best Local Similarity 71.4%; Pred. No. 0.46;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDKNPMPGA 14  
|||||:|:| |  
Db 857 GVKGERGSPGPGA 870

RESULT 21

CG8075  
collagen alpha 1(III) chain - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text\_change 07-May-1999  
C:Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946  
R:Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wächter, E.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
A:Title: The covalent structure of calf skin type III collagen. I. The amino acid seq  
A:Reference number: A02862; MUID:80026026  
A:Accession: A02862  
A:Molecule type: protein  
A:Residues: 1-242 <FIE>  
R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
A:Title: The covalent structure of calf skin type III collagen. II. The amino acid se  
A:Reference number: A38001; MUID:80026027  
A:Accession: A38001  
A:Molecule type: protein  
A:Residues: 243-422 <DEW1>  
R:Bentz, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
A:Title: The covalent structure of calf skin type III collagen. III. The amino acid s  
A:Reference number: A38002; MUID:80026028  
A:Accession: A38002  
A:Molecule type: protein  
A:Residues: 423-571 <BEN>  
R:Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
A:Title: The covalent structure of calf skin type III collagen. IV. The amino acid se  
A:Reference number: A38003; MUID:80026029  
A:Accession: A38003  
A:Molecule type: protein  
A:Residues: 572-808 <LAN>  
R:Dewes, H.; Fietzek, P.P.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
A:Title: The covalent structure of calf skin type III collagen. V. The amino acid seq  
A:Reference number: A38004; MUID:80026030  
A:Accession: A38004  
A:Molecule type: protein  
A:Residues: 809-947 <DEW2>  
R:Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.  
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
A:Title: The covalent structure of calf skin type III collagen. VI. The amino acid se  
A:Reference number: A38005; MUID:80026031  
A:Accession: A38005  
A:Molecule type: protein  
A:Residues: 948-1049 <ALU>  
A:Experimental source: skin  
R:Henkel, W.  
Biochem. J. 318, 497-503, 1996  
A:Title: Cross-link analysis of the C-telopeptide domain from type III collagen.  
A:Reference number: S71946; MUID:96404897  
A:Accession: S71946  
A:Molecule type: protein  
A:Residues: 87-106;1017-1029;1037-1049 <HEN>  
C:Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) ar  
C:Comment: The type III collagen molecule is a trimer of identical chains, linked to  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
F:1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>  
F:1-14/Region: amino-terminal nonhelical telopeptide  
F:15-1040/Region: helical

F:587-589/Region: cell attachment (R-G-D) motif  
F:752-754/Region: cell attachment (R-G-D) motif  
F:755-877/Region: cell attachment (R-G-D) motif  
F:878-880/Region: cell attachment (R-G-D) motif  
F:935-937/Region: cell attachment (R-G-D) motif  
F:1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
F:107-950/Modified site: lysine (Lys) #status experimental  
F:107/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F:1040/1041/Disulfide bonds: Interchain #status predicted

Query Match 65.2%; Score 60; DB 1; Length 1049;  
Best Local Similarity 71.4%; Pred. No. 0.54;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKGDKNPGMPCA 14  
Db 702 GVKGERGSPGPGA 715

## RESULT 22

CGHUTL  
collagen alpha 1(III) chain precursor - human  
N:Alternate names: procollagen alpha 1(III) chain  
C:Species: Homo sapiens (man)  
C>Date: 24-Apr-1984 #sequence, revision 01-Sep-1995 #text, change 21-Jul-2000  
C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90  
R:Prockop, D.J.  
Submitted to the EMBL Data Library, February 1989  
A:Reference number: S05272  
A:Accession: S05272  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1240, 'V', 1242-1466 <PRC>  
A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058  
R:Ala-Kokko, L.; Kontunen, S.; Baldwin, C.T.; Kallvinnem, H.; Prockop, D.J.  
Biochem. J. 260, 509-516, 1989  
A:Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human  
ence.  
A:Reference number: S04642; MUID:89350838  
A:Accession: S04642  
A:Molecule type: mRNA  
A:Residues: 1-1196 <ALA>  
A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058  
R:Note: the complete sequence is not shown  
R:Benson-Chanda, V.; Su, M.W.; Well, D.; Chu, M.L.; Ramirez, F.  
Gene 78, 253-265, 1989  
A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene  
A:Reference number: PE0011; MUID:89378752  
A:Accession: PE0011  
A:Molecule type: DNA  
A:Residues: 1-176 <BEN>  
A:Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814  
R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.  
Nucleic Acids Res. 16, 7201, 1988  
A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre  
A:Reference number: S01726; MUID:88303360  
A:Accession: S01726  
A:Molecule type: mRNA  
A:Residues: 1-170 <ROM>  
A:Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061  
R:Note: the authors translated the codon CAG for residue 154 as His  
R:Janeczko, R.A.; Ramirez, F.  
Nucleic Acids Res. 17, 6742, 1989  
A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.  
A:Reference number: S04887; MUID:89386015  
A:Accession: S04887  
A:Molecule type: mRNA  
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'V', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,  
A:Cross-references: EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:g930045  
R:Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 16, 1158-1164, 1977

A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide pept  
A:Reference number: A90399; MUID:77134734  
A:Accession: A90399  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
A:Experimental source: liver  
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-gal  
R:Seyer, J.M.  
submitted to the Atlas, December 1977  
A:Reference number: A94562  
A:Accession: A94562  
A:Molecule type: protein  
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A:Experimental source: liver  
A:Note: author submitted corrections to A90399  
R:Milwicz, D.M.; Wiltz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A:Title: Parental somatic and germ-line mosaicism for a multigene deletion with unusu  
spring.  
A:Reference number: I51868; MUID:93304430  
A:Accession: I51868  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 186-194 <MTL>  
A:Cross-references: GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:g4261637  
R:Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the C  
A:Reference number: S59511; MUID:96067614  
A:Accession: S59511  
A:Molecule type: mRNA  
A:Residues: 302-423 <CHD>  
A:Cross-references: GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:g1195577  
R:Seyer, J.M.; Kang, A.H.  
Biochemistry 17, 3404-3411, 1978  
A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBR  
A:Reference number: A90414; MUID:79000343  
A:Accession: A90414  
A:Molecule type: protein  
A:Residues: 399-675, 'N', 677-727 <SEY3>  
A:Experimental source: liver  
R:Lee, B.; Vitale, E.; Superl-Furga, A.; Stelmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5256-5259, 1991  
A:Title: G to T transversion at position +5 of a splice donor site causes skipping of  
A:Reference number: I55349; MUID:91161621  
A:Accession: I55349  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 537-605 <LEE>  
A:Cross-references: GB:M59312; NID:g180815; PIDN:AAA52041.1; PID:g180816  
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1980  
A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-C85 from  
A:Reference number: A90438; MUID:80198282  
A:Accession: A90438  
A:Molecule type: protein  
A:Residues: 728-855, 'A', 857-964 <SEY4>  
A:Experimental source: liver  
R:Lee, W.G.; Chiodo, A.A.; Lemande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; G  
J. Biol. Chem. 265, 17070-17077, 1990  
A:Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping  
A:Reference number: A38303; MUID:91009133  
A:Accession: A38303  
A:Molecule type: mRNA  
A:Residues: 861-1015 <COL>  
A:Cross-references: GB:J05617; GB:M55603; GB:M59227; NID:g180878; PIDN:AAB59383.1; PI  
A:Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos  
R:Mankoo, B.S.; Dalglish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A:Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A:Reference number: S02119; MUID:88189837  
A:Accession: S02119  
A:Status: translation not shown

A:Molecule type: mRNA  
A:Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>  
A:Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID:g30054  
R:Sever, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A:Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from cy  
A:Reference number: A90446; MUID:81208139  
A:Accession: A90446  
A:Molecule type: Protein  
A:Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AR', 1155, 'S', 1157-  
A:Experimental source: liver  
R:Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye  
Nucleic Acids Res. 12, 9383-9394, 1984  
A:Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage  
A:Reference number: A93551; MUID:85087944  
A:Accession: A93551  
A:Molecule type: mRNA  
A:Residues: 1065-1155, 'P', 1157-1466 <LOI>  
A:Cross-references: EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CAA25821.1  
R:Miskulin, M.; Dalgleish, R.; Klueber-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant  
Biochemistry 25, 1408-1413, 1986  
A:Title: Human type III collagen gene expression is coordinately modulated with the type  
A:Reference number: 152393; MUID:86187804  
A:Accession: 152393  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1161-1200 <MIS>  
A:Cross-references: GB:M13146; NID:g180415; PIDN:AA52003.1; PID:g180416  
R:Emanuel, B.S.; Canizazaro, L.A.; Sever, J.M.; Myers, J.C.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
A:Reference number: 159025; MUID:85216505  
A:Accession: 179359  
A>Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1165-1196 <EMIA>  
A:Cross-references: GB:M11134; NID:g180417; PIDN:AA52004.1; PID:g180418  
R:Chu, M.L.; Wall, D.; de Wet, W.; Bernard, M.; Sipola, M.; Ramirez, F.  
J. Biol. Chem. 260, 4357-4363, 1985  
A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F  
A:Reference number: A92516; MUID:85157600  
A:Accession: A92516  
A:Molecule type: DNA  
A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHD>  
A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB  
A:Experimental source: liver  
A:Note: the authors translated the codon TTC for residue 1057 as Tyr: the codons given f  
ation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C  
C:Genetics:  
A:Gene: GDB:COL3A1  
A:Cross-references: GDB:118729; OMIM:120180  
A:Map position: 2q31-q31  
A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3  
A:Note: The list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
er of their length, is formed with desmosine cross-links made from lysine and allysine r  
C:Function:  
A:Description: structural component of extracellular fibrous polymer that maintains inte  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>  
F:31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>  
F:154-167/Region: amino-terminal nonhelical telopeptide  
F:168-1196/Region: helical  
F:1091-1093/Region: cell attachment (R-G-D) motif  
F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>  
F:1228-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F:124/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status pred  
F:161,1212/Modified site: allysine (Lys) #status predicted  
F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:263/Binding site: carboxylate (Lys) (covalent) #status experimental  
F:588,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental  
F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted  
Query Match 65.2%; Score 60; DB 1; Length 1466;  
Best Local Similarity 78.6%; Pred. NO. 0.76;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GVKGDKGNFGMGCA 14  
DB 738 GPKGDKGGERGCA 751  
RESULT 23  
A61262  
collagen alpha 1(XVII) chain - human (fragment)  
N:Alternate names: bullous pemphigoid 180k autoantigen BPAG2; bullous pemphigoid anti  
C:Species: Homo sapiens (man)  
C:Date: 12-May-1994 #sequence\_revision 12-Jul-1996 #text\_change 21-Jul-2000  
C:Accession: 156325; 155345; A61262  
R:Giudice, G.J.; Emery, D.J.; Diaz, L.A.  
J. Invest. Dermatol. 99, 243-250, 1992  
A:Title: Cloning and primary structural analysis of the bullous pemphigoid autoantige  
A:Reference number: 156325; MUID:92381323  
A:Accession: 156325  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1532 <RES>  
A:Cross-references: GB:M01669; NID:g179516; PIDN:AA35605.1; PID:g179517  
R:Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mettel, M.G.; Chu, M.L.; Uitto,  
J. Biol. Chem. 266, 24064-24069, 1991  
A:Title: Genomic organization of collagenous domains and chromosomal assignment of hu  
A:Reference number: 155345; MUID:92084712  
A:Accession: 155345  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 543-890, 'P' <RE2>  
A:Cross-references: GB:M63730; NID:g179520; PIDN:AA51839.1; PID:g179521  
R:Giudice, G.J.; Squigera, H.L.; Elias, P.M.; Diaz, L.A.  
J. Clin. Invest. 87, 734-738, 1991  
A:Title: Identification of two collagen domains within the bullous pemphigoid autoant  
A:Reference number: A61262; MUID:91123476  
A:Accession: A61262  
A:Molecule type: mRNA  
A:Residues: 543-890, 'P' <GID>  
C:Genetics:  
A:Gene: GDB:COL17A1; BPAG2; BP180  
A:Cross-references: GDB:131396; OMIM:113811  
A:Map position: 10q24.3-10q24.3  
Query Match 65.2%; Score 60; DB 2; Length 1532;  
Best Local Similarity 78.6%; Pred. NO. 0.79;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKGDKGNFGMGCA 14  
DB 995 GPKGDKGGERGCA 1008  
RESULT 24  
t13990  
collagen type IV alpha 2 - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C:Accession: t13990  
R:Yaoshunshikul, S.; Davis, W.J.; Cramer, G.; Kimbrell, D.A.; Dearolf, C.R.  
submitted to the EMBL Data Library, July 1996  
A:Description: Viking; identification and characterization of a novel type IV collage









Fri Nov 1 13:09:20 2002

us-09-529-691a-1.rpr

Page 17

Job time : 15.5 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:16:06 ; Search time 7 seconds

(without alignments)  
82,970 Million cell updates/sec

Title: US-09-529-691A-1

Perfect score: 92  
Sequence: 1 GVKGDKGNPGWGPAP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	1669	1 CA1A_HUMAN	P02462 homo sapien
2	78	84.8	1669	1 CA1A_MOUSE	P02463 mus musculu
3	69	75.0	1685	1 CA54_HUMAN	P29400 homo sapien
4	66	71.7	754	1 CA54_CANFA	Q28247 canis famill
5	65	70.7	1707	1 CA24_MOUSE	P08172 mus musculu
6	64	69.6	1712	1 CA24_HUMAN	P08572 homo sapien
7	63	68.5	2944	1 CA17_HUMAN	P00388 homo sapien
8	62	67.4	1670	1 CA34_HUMAN	P01955 homo sapien
9	61	66.3	339	1 COA5_BPPRD	P22536 bacterioph
10	61	66.3	1603	1 CA1F_HUMAN	Q07092 homo sapien
11	61	66.3	1690	1 CA44_HUMAN	P53420 homo sapien
12	61	66.3	1763	1 CA24_ASCSU	P27333 ascaris suu
13	60	65.2	369	1 PSPD_BOVIN	P35246 bos taurus
14	60	65.2	623	1 CA44_RABIT	P55787 oryctolagus
15	60	65.2	636	1 CA13_RAT	P13941 ratus norv
16	60	65.2	1049	1 CA13_BOVIN	P04258 bos taurus
17	60	65.2	1262	1 CA13_CHICK	P12105 gallus gall
18	60	65.2	1466	1 CA13_HUMAN	P02461 homo sapien
19	59	64.1	471	1 CA14_BOVIN	Q28044 bos taurus
20	59	64.1	1758	1 CA14_CAEEL	P17139 caenorthabd1
21	58	63.0	285	1 YRP2_CAEEL	Q09578 caenorthabd1
22	58	63.0	1516	1 CA1H_HUMAN	P39060 homo sapien
23	58	63.0	1758	1 CA24_CAEEL	P17140 caenorthabd1
24	57	62.0	298	1 CC34_CAEEL	P34687 caenorthabd1
25	57	62.0	483	1 MRCO_MESAU	P34687 mesocricetu
26	57	62.0	1356	1 CA21_ONCMY	Q93484 oncorhynch
27	56	60.9	518	1 MTCO_MOUSE	Q60754 mus musculu
28	56	60.9	1143	1 CA1I_HUMAN	Q14993 homo sapien
29	56	60.9	1464	1 CA13_MOUSE	P08121 mus musculu
30	55.5	60.3	458	1 MSRE_MOUSE	P30247 mus musculu
31	55	59.8	245	1 C1OC_HUMAN	P02747 homo sapien
32	55	59.8	266	1 YXMK_CAEEL	Q21184 caenorthabd1
33	55	59.8	371	1 CONG_BOVIN	P23805 bos taurus

34	55	59.8	680	1 CA1A_MOUSE	O05306 mus musculu
35	55	59.8	744	1 CA1B_HUMAN	P27658 homo sapien
36	55	59.8	921	1 CA1D_MOUSE	Q05722 mus musculu
37	55	59.8	963	1 YQ36_CAEEL	Q09457 caenorthabd1
38	55	59.8	1650	1 CA2B_MOUSE	Q64739 mus musculu
39	55	59.8	1678	1 CA64_HUMAN	Q14031 homo sapien
40	55	59.8	1736	1 CA2B_HUMAN	P13942 homo sapien
41	54	58.7	316	1 CC12_CAEEL	P20630 caenorthabd1
42	54	58.7	316	1 CC13_CAEEL	P20631 caenorthabd1
43	54	58.7	333	1 CA29_CHICK	P12108 gallus gall
44	54	58.7	375	1 PSPD_HUMAN	P35247 homo sapien
45	54	58.7	635	1 CA2B_HUMAN	P25067 homo sapien

## ALIGNMENTS

RESULT 1	ID	CA1A_HUMAN	STANDARD:	PRT:	1669 AA.
AC	P02462:	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Collagen alpha 1(IV) chain precursor.				
GN	COL4A1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid:9606;				
RN	[1]	SEQUENCE FROM N.A.			
RP	MDLINE-89340433; PubMed-2701944;				
RX	Solinen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;				
RT	"Structural organization of the gene for the alpha 1 chain of human				
RT	type IV collagen.";				
RL	J. Biol. Chem. 264:13565-13571(1989).				
RN	[2]	SEQUENCE OF 46-1257 FROM N.A.			
RP	TISSUE-Placenta;				
RC	MDLINE-88083584; PubMed-3691802;				
RX	Solinen R., Haka-Risku T., Prockop D.J., Tryggvason K.;				
RT	"Complete primary structure of the alpha 1-chain of human basement				
RT	membrane (type IV) collagen.";				
RL	FEBS Lett. 225:188-194(1987).				
RN	[3]	SEQUENCE OF 1-943 FROM N.A.			
RP	TISSUE-Placenta;				
RC	MDLINE-88029471; PubMed-3311751;				
RX	Brazel D., Oberbauer I., Dieringer H., Babel W., Gianville R.W.,				
RA	Deutzmann R., Kuehn K.;				
RT	"Completion of the amino acid sequence of the alpha 1 chain of human				
RT	basement membrane collagen (type IV) reveals 21 non-triplet				
RT	interruptions located within the collagenous domain.";				
RL	Eur. J. Biochem. 168:529-536(1987).				
RN	[4]	SEQUENCE OF 28-243.			
RP	MDLINE-86004708; PubMed-4043082;				
RX	Gianville R.W., Qian R.O., Siebold B., Ristell J., Kuehn K.;				
RA	"Amino acid sequence of the N-terminal aggregation and cross-linking				
RT	region (7S domain) of the alpha 1 (IV) chain of human basement				
RT	membrane collagen.";				
RL	Eur. J. Biochem. 152:213-219(1985).				
RN	[5]	SEQUENCE OF 534-1447.			
RP	MDLINE-85003629; PubMed-6434307;				
RX	Babel W., Gianville R.W.;				
RA	"Structure of human-basement-membrane (type IV) collagen. Complete				
RT	amino-acid sequence of a 914-residue-long pepsin fragment from the				
RT	alpha 1(IV) chain.";				
RL	Eur. J. Biochem. 143:545-556(1984).				
RN	[6]	SEQUENCE OF 1256-1669 FROM N.A.			

RX MEDLINE-85207819; PubMed-2581969;  
 RA PhilaJanemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,  
 Cheung M.-C., Prockop D.J., Boyd C.D.;  
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV  
 RT procollagen reveal an unusual homology of amino acid sequences in two  
 RT halves of the carboxyl-terminal domain."  
 RL J. Biol. Chem. 260:7681-7687(1985).  
 RN [17]  
 RP SEQUENCE OF 1259-1669 FROM N.A.  
 RX MEDLINE-85216555; PubMed-2582422;  
 RA Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,  
 RT "Restricted homology between human alpha 1 type IV and other  
 RT procollagen chains."  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).  
 RN [18]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE-89034231; PubMed-3182844;  
 RA Solinlen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;  
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV  
 RT collagen are divergently encoded on opposite DNA strands and have an  
 RT overlapping promoter region."  
 RL J. Biol. Chem. 263:17217-17220(1988).  
 RN [9]  
 RP SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.  
 RX TISSUE-Placenta;  
 RC MEDLINE-89005112; PubMed-2844531;  
 RA Siebold B., Deutmann R., Kuehn K.;  
 RT "The arrangement of intra- and intermolecular disulfide bonds in the  
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain  
 RT of basement-membrane type IV collagen."  
 RL Eur. J. Biochem. 176:617-624(1988).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCAN AND ENACTIN/  
 CC NIDOGN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PPM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDATE.  
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CISTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; M26576; AAA53098.1; JOINED.  
 DR EMBL; J04217; AAA53098.1; JOINED.  
 DR EMBL; M26550; AAA53098.1; JOINED.  
 DR EMBL; M26540; AAA53098.1; JOINED.  
 DR EMBL; M26542; AAA53098.1; JOINED.  
 DR EMBL; M26543; AAA53098.1; JOINED.  
 DR EMBL; M26544; AAA53098.1; JOINED.  
 DR EMBL; M26545; AAA53098.1; JOINED.  
 DR EMBL; M26546; AAA53098.1; JOINED.  
 DR EMBL; M26547; AAA53098.1; JOINED.  
 DR EMBL; M26537; AAA53098.1; JOINED.  
 DR EMBL; M26538; AAA53098.1; JOINED.

DR EMBL; M26548; AAA53098.1; JOINED.  
 DR EMBL; M26549; AAA53098.1; JOINED.  
 DR EMBL; M26551; AAA53098.1; JOINED.  
 DR EMBL; M26552; AAA53098.1; JOINED.  
 DR EMBL; M26553; AAA53098.1; JOINED.  
 DR EMBL; M26554; AAA53098.1; JOINED.  
 DR EMBL; M26555; AAA53098.1; JOINED.  
 DR EMBL; M26556; AAA53098.1; JOINED.  
 DR EMBL; M26557; AAA53098.1; JOINED.  
 DR EMBL; M26559; AAA53098.1; JOINED.  
 DR EMBL; M26560; AAA53098.1; JOINED.  
 DR EMBL; M26561; AAA53098.1; JOINED.  
 DR EMBL; M26562; AAA53098.1; JOINED.  
 DR EMBL; M26563; AAA53098.1; JOINED.  
 DR EMBL; M26564; AAA53098.1; JOINED.  
 DR EMBL; M26565; AAA53098.1; JOINED.  
 DR EMBL; M26566; AAA53098.1; JOINED.  
 DR EMBL; M26567; AAA53098.1; JOINED.  
 DR EMBL; M26568; AAA53098.1; JOINED.  
 DR EMBL; M26569; AAA53098.1; JOINED.  
 DR EMBL; M26570; AAA53098.1; JOINED.  
 DR EMBL; M26571; AAA53098.1; JOINED.  
 DR EMBL; M26572; AAA53098.1; JOINED.  
 DR EMBL; M26573; AAA53098.1; JOINED.  
 DR EMBL; M26574; AAA53098.1; JOINED.  
 DR EMBL; M26575; AAA53098.1; JOINED.  
 DR EMBL; M26576; AAA53098.1; JOINED.  
 DR EMBL; Y00706; CAA8698.1; JOINED.  
 DR EMBL; X05561; CAA29075.1; JOINED.  
 DR EMBL; M10940; AAA52006.1; JOINED.  
 DR EMBL; M11315; AAA52042.1; JOINED.  
 DR PIR; S16876; CGH04B.  
 DR MIM; 120130; IPR001442; C4.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.  
 DR ProDom; PD003923; C4; 2.  
 DR SMART; SM00111; C4; 2.  
 DR KEGG; Extracellular matrix; Connective tissue; Basement membrane;  
 DR KEGG; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 27  
 FT PROPEP 28 172  
 FT CHAIN 173 1669  
 FT DOMAIN 173 1440  
 FT CARBOHYD 1441 1669  
 FT DISULFID 126 126  
 FT DISULFID 1460 1551  
 FT DISULFID 1493 1548  
 FT DISULFID 1505 1511  
 FT DISULFID 1570 1665  
 FT DISULFID 1604 1662  
 FT DISULFID 1616 1622  
 FT DISULFID 237 238  
 FT CONFLICT 241 241  
 FT CONFLICT 319 319  
 FT CONFLICT 719 719  
 FT CONFLICT 837 837  
 FT CONFLICT 842 842  
 FT CONFLICT 896 896  
 FT CONFLICT 904 904  
 FT CONFLICT 914 914  
 FT CONFLICT 998 998  
 FT CONFLICT 1010 1010  
 FT CONFLICT 1012 1012  
 FT CONFLICT 1358 1358  
 FT SEQUENCE 1669 AA; 160611 MW; 3BBBA6DFFB9B8A84 CRC64;  
 SG -> KE (IN REF. 4).  
 G -> K (IN REF. 4).  
 Q -> A (IN REF. 3).  
 N -> D (IN REF. 5).  
 D -> Y (IN REF. 5).  
 K -> P (IN REF. 5).  
 V -> W (IN REF. 5).  
 E -> Q (IN REF. 5).  
 S -> K (IN REF. 5).  
 S -> P (IN REF. 5).  
 K -> K (IN REF. 5).  
 S -> K (IN REF. 5).  
 E -> Q (IN REF. 5).  
 OR 1662.  
 OR 1665.  
 OR 1548.  
 OR 1551.  
 OR 1662.  
 OR 1665.

Query Match 100.0%; Score 92; DB 1; Length 1669;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GYKDGKNGPGAP 15  
Db 1263 GYKDGKNGPGAP 1277

RESULT 2  
CA14\_MOUSE STANDARD; PRT; 1669 AA.  
AC P02463;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.  
GN COL4A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89197932; PubMed=2703490;  
RA Muthukumar G., Blumberg B., Kurkinen M.;  
RT "The complete primary structure for the alpha 1-chain of mouse  
RT collagen IV. Differential evolution of collagen IV domains.";  
RL J. Biol. Chem. 264:6310-6317(1989).  
RN [2]  
RP SEQUENCE OF 1-1154 FROM N.A.  
RX MEDLINE=88112221; PubMed=3338568;  
RA Wood L., Theriault N., Vogel G.;  
RT "CDNA clones completing the nucleotide and derived amino acid  
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen  
RT from mouse.";  
RL FEBS Lett. 227:5-8(1988).  
RN [3]  
RP SEQUENCE OF 1149-1424 FROM N.A.  
RX MEDLINE=86301886; PubMed=3755692;  
RA Math P., Laurent M., Horn E., Sobel M.E., Zon G., Vogel G.;  
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a  
RT synthetic oligodeoxynucleotide.";  
RL Gene 43:301-304(1986).  
RN [4]  
RP SEQUENCE OF 1276-1669 FROM N.A.  
RX MEDLINE=85127033; PubMed=2578961;  
RA Oberhaeumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
RA Vogel G., Voss T., Stebold B., Glasville R.W., Kuhn K.;  
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of  
RT the alpha 1(IV) chain of basement membrane collagen as derived from  
RT complementary DNA.";  
RL Eur. J. Biochem. 147:217-224(1985).  
RN [5]  
RP SEQUENCE OF 1441-1669 FROM N.A.  
RX MEDLINE=87250460; PubMed=3597383;  
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
RA Saus J., Pihlajaniemi T.;  
RT "Extensive homology between the carboxyl-terminal peptides of mouse  
RT alpha 1(IV) and alpha 2(IV) collagens.";  
RL J. Biol. Chem. 262:8496-8499(1987).  
RN [6]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=86196099; PubMed=3009468;  
RA Sakurai Y., Sullivan M., Yamada Y.;  
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar  
RT collagen genes.";  
RL J. Biol. Chem. 261:6654-6657(1986).  
RN [7]  
RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89066738; PubMed=3198626;  
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogel G.;  
RT "Head-to-head arrangement of murine type IV collagen genes.";  
RL J. Biol. Chem. 263:19274-19277(1988).  
RN [8]

RP SEQUENCE OF 1-28 FROM N.A.  
RX MEDLINE=89071759; PubMed=3200851;  
RA Burbelo P.D., Martin G.R., Yamada Y.;  
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a  
RT bidirectional promoter and a shared enhancer.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).  
RN [9]  
RP SEQUENCE OF 1-129 FROM N.A.  
RX MEDLINE=88243724; PubMed=3379041;  
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;  
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)  
RT collagen chain and the corresponding region of the gene.";  
RL J. Biol. Chem. 263:8706-8709(1988).  
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
CC NIDOGEN.  
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
CC TRIPLE-HELICAL 7S DOMAIN.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
CC IV COLLAGENS.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J03758; AAA37439.1; -  
DR EMBL: M23333; AAA51625.1; -  
DR EMBL: J04694; AAA50292.1; -  
DR EMBL: X06777; CAA29946.1; -  
DR EMBL: X02201; CAA26132.1; -  
DR EMBL: M15832; AAA37340.1; -  
DR EMBL: M14042; AAA37342.1; -  
DR EMBL: M12879; AAA37343.1; -  
DR EMBL: M13024; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: M13025; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: M13026; AAA37344.1; -  
DR EMBL: M13027; AAA37345.1; -  
DR EMBL: M13043; AAA37346.1; -  
DR EMBL: J04448; AAA37437.1; -  
DR PIR: A33525; CGMS4B.  
DR MGI: 88454; Col4a1.  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR Pfam: PF01391; Collagen; 21.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Extracellular matrix; Connective tissue; Basement membrane;  
KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.  
FT SIGNAL 1 27  
FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).  
FT DISULFID 1460 1551 OR 1348 (BY SIMILARITY).  
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
FT DISULFID 1505 1511 BY SIMILARITY.  
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).  
FT

FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).  
 FT DISULFID 1616 1622 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 26 26 A -> P (IN REF. 2).  
 FT CONFLICT 186 186 S -> L (IN REF. 2).  
 FT CONFLICT 319 319 Q -> S (IN REF. 2).  
 FT CONFLICT 369 369 Q -> L (IN REF. 2).  
 FT CONFLICT 403 403 L -> F (IN REF. 2).  
 FT CONFLICT 481 481 P -> L (IN REF. 2).  
 FT CONFLICT 493 493 Q -> H (IN REF. 2).  
 FT CONFLICT 712 712 S -> I (IN REF. 2).  
 FT CONFLICT 813 813 E -> Q (IN REF. 2).  
 FT CONFLICT 982 982 Q -> H (IN REF. 2).  
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).  
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match 84.8%; Score 78; DB 1; Length 1669;  
 Best Local Similarity 86.7%; Pred. NO. 0.00065;  
 Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

DY 1 GVGDKGNPMPGAP 15  
 1 ||||| |||||  
 Db 1263 GPKGDKGNMPGAP 1277

RESULT 3  
 ID CA54\_HUMAN STANDARD: PRT; 1685 AA.  
 AC P29400; O16126; O16006;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 5(IV) chain precursor.  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCI; TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94165049; PubMed=8120014;  
 RA Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RL J. Biol. Chem. 269:6608-6614(1994).  
 RN [2]  
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.  
 RC TISSUE=Kidney;  
 RX MEDLINE=92316923; PubMed=1355287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen  
 chain and identification of a single-base mutation in exon 23  
 converting glycine 521 in the collagenous domain to cysteine in an  
 Alport syndrome patient.";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 RN [3]  
 RP SEQUENCE OF 85-1685 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90337990; PubMed=2380186;  
 RA Pihlajaniemi T., Pohjolaenen E.R., Myers J.C.;  
 RT "Complete primary structure of the triple-helical region and the  
 carboxyl-terminal domain of a new type IV collagen chain, alpha  
 5(IV).";  
 RL J. Biol. Chem. 265:13758-13766(1990).  
 RN [4]  
 RP SEQUENCE OF 924-1685 FROM N.A.  
 RX MEDLINE=91168491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RT "Characterization of the 3' half of the human type IV collagen alpha  
 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 RN [5]  
 RP SEQUENCE OF 914-1685 FROM N.A.  
 RX MEDLINE=90160375; PubMed=1689491;

RA Hostikka S.L., Eddy R.L., Byers M.G., Hoehlytae M., Shows T.B.,  
 RA Tryggvason K.;  
 RT "Identification of a distinct type IV collagen alpha chain with  
 restricted kidney distribution and assignment of its gene to the  
 locus of X chromosome-linked Alport syndrome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6]  
 RP SEQUENCE OF 1442-1471 FROM N.A.  
 RX MEDLINE=90252791; PubMed=2339699;  
 RA Myers J.C., Jones T.A., Pihlajaniemi E.R., Kadri A.S., Goddard A.D.,  
 RA Sheer D., Solomon E., Pihlajaniemi T.;  
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene  
 to the region of the X chromosome containing the Alport syndrome  
 locus.";  
 RL Am. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,  
 RA Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN [8]  
 RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).  
 RX MEDLINE=94133540; PubMed=8301933;  
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,  
 RA Cassiman J.-J., Marynen P.;  
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood  
 cells: a complex mutation in the COL4A5 gene of an Alport patient  
 deletes the NCI domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE=97338662; PubMed=9195222;  
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "Type clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [10]  
 RP VARIANT AS SER-1564.  
 RX MEDLINE=91169492; PubMed=1672282;  
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Acklin C.L.,  
 RA Tryggvason K.;  
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a  
 conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]  
 RP VARIANT AS ARG-325.  
 RX MEDLINE=92303559; PubMed=1376965;  
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,  
 RA Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5  
 (IV) chain associated with X-linked Alport syndrome: characterization  
 of the mutation by direct sequencing of PCR-amplified lymphoblast  
 cDNA fragments.";  
 RL Am. J. Hum. Genet. 51:135-142(1992).  
 RN [12]  
 RP VARIANT AS GLU-325.  
 RX MEDLINE=93244772; PubMed=1363780;  
 RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,  
 RA Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to  
 glutamic acid in Alport syndrome.";  
 RL Hum. Mol. Genet. 1:127-129(1992).  
 RN [13]  
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.  
 RX MEDLINE=94010948; PubMed=8406498;  
 RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,  
 RA Tryggvason K., Hagsma-Schouten W.A.G., Roodvoets A.P., Rascher W.,  
 RA van Oost B.A., Smeets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of  
 patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]  
 RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872  
 AND C-1241.

RA MEDLINE-95322976; PubMed-7599631;  
 RA Boye E., Fliether F., Zhou J., Tryggvason K., Bobrow M., Harris A.;  
 RT "Detection of 12 novel mutations in the collagenous domain of the  
 RT COL4A5 gene in Alport syndrome patients."; Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE-96213750; PubMed-8651292;  
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,  
 RA Denison J.C., Fain P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 RT common in the western United States."; Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANTS AS  
 RX MEDLINE-96213754; PubMed-8651296;  
 RA Renieri A., Brutelli M., Galli L., Zanelli P., Neri T.M., Rossetti S.,  
 RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,  
 RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatelli P.F.,  
 RA Sevi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51  
 RT exons of the COL4A5 gene."; Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND  
 RP MET-1428.  
 RX MEDLINE-97094179; PubMed-8940267;  
 RA Knedelmann B., Brellat C., Forestier L., Arrondel C., Jacassier D.,  
 RA Glatas I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,  
 RA Gubler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 RT syndrome."; Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-1498.  
 RX MEDLINE-96233932; PubMed-8829632;  
 RA Tverekaya S., Bobryna V., Tsalykova F., Ignatova M.,  
 RA Krasnopol'skaya X., Evgrafov O.;  
 RT "Substitution of A1498D in noncollagen domain of  $\alpha 5(\text{IV})$  collagen  
 RT chain associated with adult-onset X-linked Alport syndrome."; Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-1677.  
 RX MEDLINE-97295089; PubMed-9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 RT syndrome and COL4A5 R1677Q."; Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.  
 RX MEDLINE-98112435; PubMed-9452056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Pignatelli G.F., Galli L., Brutelli M., Renieri A., Mingarelli R.,  
 RA Trivelli A., Pinciaroli A.R., Regalado M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 RT Alport syndrome."; Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANTS AS V-420; A56-P-P-458 DEL; D-573; D-624; D-635; 802-G-P-807  
 RP DEL; R-669; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357  
 RP AND R-1649.  
 RX MEDLINE-99063529; PubMed-9848783;  
 RA Martin P., Heiskari N., Zhou J., Ielonen A., Tuomelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Strykarsdotter U., Neumann H.,  
 RA Sprigatje J., Shows T.B., Pelterson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RT Alport syndrome using PCR and direct DNA sequencing."; J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.  
 RX MEDLINE-20030197; PubMed-10561141;  
 RA Inoue Y., Nishio H., Shitakawa T., Nakanishi K., Nakamura H.,

RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;  
 RT "Detection of mutations in the COL4A5 gene in over 90% of male  
 RT patients with X-linked Alport's syndrome by RT-PCR and direct  
 RT sequencing."; Am. J. Kidney Dis. 34:854-862(1999).  
 RN [23]  
 RP VARIANT AS ARG-822.  
 RX MEDLINE-20025011; PubMed-10563487;  
 Query Match 75.0%; Score 69; DB 1; Length 1685;  
 Best Local Similarity 73.3%; Pred. No. 0.014;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GYKDGKNGPGMPGAP 15  
 Db 1279 GIKGKGNPGPGPLP 1293  
 |||:||||| |||  
 RESULT 4  
 ID CA54\_CANFA STANDARD; PRT; 754 AA.  
 AC Q28247;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Collagen alpha 5(IV) chain (Fragment).  
 GN COL4A5.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_Taxid=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SAWOTED; TISSUE=Kidney;  
 RX MEDLINE-94224868; PubMed-8171024;  
 RA Zheng K., Thorner P.S., Mariano P., Bauman R., McInnes R.R.;  
 RT "Canine x chromosome-linked hereditary nephritis: a genetic model for  
 RT human X-linked hereditary nephritis resulting from a single base  
 RT mutation in the gene encoding the alpha 5 chain of collagen type  
 RT IV."; Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).  
 RL CC  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/  
 CC NIDOGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF  
 CC CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO  
 CC THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED  
 CC BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: U07888; AAB60258.1; -.
DR HSSP: Q04656; 2AM0.
DR InterPro: IPR001442; C4.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 8.
DR ProDom: PD003923; C4; 2.
DR SMART: SM00111; C4; 2.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1
FT DOMAIN 1 530 TRIPLE-HELICAL REGION.
FT DOMAIN 531 >754 NONHELICAL REGION (NC1).
FT DISULFD 552 643 OR 640 (BY SIMILARITY).
FT DISULFD 585 640 OR 643 (BY SIMILARITY).
FT DISULFD 597 603 BY SIMILARITY.
FT DISULFD 662 603 OR 754 (BY SIMILARITY).
FT DISULFD 696 754 BY SIMILARITY.
FT DISULFD 708 714 BY SIMILARITY.
FT NON_TER 754
SQ SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 71.7%; Score 66; DB 1; Length 754;
Best Local Similarity 66.7%; Pred. No. 0.018;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGWP GAP 15
Db 355 GIKGERGNPGOPGP 369

RESULT 5
CA24_MOUSE STANDARD: PRT: 1707 AA.
ID CA24_MOUSE
AC P08122; Q61375;
DT 01-AUG-1988 (Rel. 08, Created)
DT 31-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197933; Pubmed=2703491;
RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
RA Philajantem T., Kurkinen M.;
RT "The complete primary structure of mouse alpha 2(IV) collagen.
RT Alignment with mouse alpha 1(IV) collagen.";
RL J. Biol. Chem. 264:6318-6324(1989).
RN [2]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89066738; Pubmed=3198626;
RA Kayes P., Wood L., Theriault N., Kurkinen M., Vogel G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [3]
RP SEQUENCE OF 970-1480 FROM N.A.
RX MEDLINE=86220192; Pubmed=3011432;
RA Schwarz U., Schuppan D., Oberhauser I., Glanville R.W.,
RA Deutzmann R., Timpl R., Kuehn K.;
RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-
RT terminal 511-residue-long triple-helical segment of the alpha 2(IV)
RL Eur. J. Biochem. 157:49-56(1986).
RN [4]
RP SEQUENCE OF 1480-1707 FROM N.A.
RX MEDLINE=87054581; Pubmed=3780963;
RA Schwarz-Magdolen U., Oberhauser I., Kuehn K.;
RT "cDNA and protein sequence of the NC1 domain of the alpha 2-chain of

```

```

RT collagen IV and its comparison with alpha 1(IV).";
RL FEMS Lett. 208:203-207(1986).
RN [5]
RP SEQUENCE OF 1481-1707 FROM N.A.
RX MEDLINE=87250460; Pubmed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Philajantem T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP SEQUENCE OF 1041-1489 FROM N.A.
RX MEDLINE=87005245; Pubmed=3758345;
RA Vogel G., Horn E., Carter J., Kayes P.S.;
RT "Proposed alignment of helical interruptions in the two subunits of
RT the basement membrane (type IV) collagen.";
RL FEBS Lett. 206:29-32(1986).
RN [7]
RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
RX MEDLINE=85296379; Pubmed=3839908;
RA Kurkinen M., Bernard M.P., Barlow D.P., Chow L.T.;
RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse
RT alpha 2(IV) collagen gene.";
RL Nature 317:177-179(1985).
RN [8]
RP SEQUENCE OF 1-60 FROM N.A.
RX MEDLINE=89071759; Pubmed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOLYCAN AND EXTRACTIN/
CC NIDOGEN.
CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -----
DR EMBL: M23334; AA51626.1; -
DR EMBL: M23333; AA51626.1; JOINED.
DR EMBL: J04695; AA50293.1; -
DR EMBL: J04448; AA37438.1; -
DR EMBL: X04647; CA28308.1; -
DR EMBL: M15833; AA37341.1; -
DR EMBL: X04410; CA22798.1; -
DR EMBL: X02896; CA26655.1; -
DR EMBL: X02897; CAB51614.1; -
DR EMBL: X02898; CA26657.1; -
DR EMBL: X02899; CA26658.1; -
DR EMBL: M23334; AA51627.1; -
DR PIR: A33526; A33526.
DR HSSP: P19972; IKVD.
DR MGI: 88455; Col4a2.

```





CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL: X05562; CAA29076.1; -  
 DR EMBL: X05610; CAA29098.1; -  
 DR EMBL: J02760; AAA58422.1; -  
 DR EMBL: M36963; AAA53099.1; -  
 DR EMBL: X12784; CAA31275.1; -  
 DR EMBL: J04217; AAA53097.1; -  
 DR PIR: A32024; A32024.  
 DR MIM: 120090; -  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 20.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25  
 FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).  
 FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.  
 FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).  
 FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).  
 FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).  
 FT DISULFID 1549 1555 BY SIMILARITY.  
 FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).  
 FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).  
 FT DISULFID 1658 1665 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...).  
 FT CONFLICT 471 471 R -> P (IN REF. 2).  
 FT CONFLICT 683 683 A -> G (IN REF. 2).  
 FT CONFLICT 1575 1575 M -> I (IN REF. 5).  
 FT CONFLICT 1663 1663 G -> H (IN REF. 9).  
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).  
 SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;  
 Query Match 69.6%; Score 64; DB 1; Length 1712;  
 Best Local Similarity 73.3%; Fred. No. 0.081;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GVKGDKGNPGAP 15  
 Db 1368 GPKGPKGDPGFGAP 1382  
 RESULT 7  
 CA17\_HUMAN STANDARD: PRT: 2944 AA.  
 ID CA17\_HUMAN Q02388; Q14054; Q16507;  
 AC Q02388; Q14054; Q16507;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC  
 DE collagen).  
 GN COL7A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94327588; PubMed=8051117;  
 RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;  
 RT "Cloning of human type VII collagen. Complete primary sequence of the  
 RT alpha 1(VII) chain and identification of intragenic polymorphisms."; J  
 RT J Biol. Chem. 269:20256-20262(1994).  
 RN [2]  
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=93338437; PubMed=1307247;  
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,  
 RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E.,  
 RA Uitto J.;  
 RT "The large non-collagenous domain (NC-1) of type VII collagen is  
 RT amino-terminal and chimeric. Homology to cartilage matrix protein,  
 RT the type III domains of fibronectin and the A domains of von  
 RT Willebrand factor."; Hum. Mol. Genet. 1:475-481(1992).  
 RN [3]  
 RP SEQUENCE OF 815-1439 FROM N.A.  
 RX MEDLINE=91334380; PubMed=1871109;  
 RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W.,  
 RA Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;  
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the  
 RT gene."; Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).  
 RN [4]  
 RP SEQUENCE OF 369-1255 FROM N.A.  
 RX MEDLINE=93107742; PubMed=1469284;  
 RA Gannon W.R., Abernethy M.L., Padilla K.M., Prisaayanh P.S.,  
 RA Cook M.E., Wright J., Brigaman R.A., Hunt S.W. III;  
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain  
 RT adhesion proteins involved in tissue-specific organization of  
 RT extracellular matrix."; J. Invest. Dermatol. 99:691-696(1992).  
 RN [5]  
 RP SEQUENCE OF 340-675 FROM N.A.  
 RX TISSUE-Keratinocytes;  
 MEDLINE=92231902; PubMed=1567409;  
 RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;  
 RT "Molecular cloning and characterization of type VII collagen cDNA."; Biochem. Biophys. Res. Commun. 183:958-963(1992).  
 RN [6]  
 RP SEQUENCE OF 2395-2944 FROM N.A.  
 RX MEDLINE=93271985; PubMed=8499916;  
 RA Greenspan D.S.;  
 RT "The carboxyl-terminal half of type VII collagen, including the non-  
 RT collagenous NC-2 domain and intron/exon organization of the  
 RT corresponding region of the COL7A1 gene."; Hum. Mol. Genet. 2:273-278(1993).  
 RN [7]  
 RP SEQUENCE OF 1-87 FROM N.A.  
 RX TISSUE=Placenta;  
 MEDLINE=94375010; PubMed=8088784;  
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,  
 RA Uitto J., Greenspan D.S.;  
 RT "Structural organization of the human type VII collagen gene (COL7A1),  
 RT composed of more exons than any previously characterized gene."; Genomics 21:169-179(1994).  
 RN [8]  
 RP REVIEW ON DEB VARIANTS.  
 RX MEDLINE=98041696; PubMed=9375848;  
 RA Jaervakallio A., Pulkkinen L., Uitto J.;  
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in  
 RT the type VII collagen gene (COL7A1)."; Hum. Mutat. 10:338-347(1997).  
 RN [9]  
 RP VARIANT RDBE LYS-2798.  
 RX MEDLINE=93291877; PubMed=8513326;  
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,  
 RA Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;  
 RT "A missense mutation in type VII collagen in two affected siblings  
 RT with recessive dystrophic epidermolysis bullosa."; Nat. Genet. 4:62-66(1993).  
 RN [10]  
 RP VARIANT DEEB SER-2040.  
 RX MEDLINE=94224777; PubMed=8170945;  
 RA Christiano A.M., Ryyanen M., Uitto J.;  
 RT "Dominant dystrophic epidermolysis bullosa: identification of a  
 RT gly--ser substitution in the triple-helical domain of type VII  
 RT collagen."; Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).

RN [11]  
 RP VARIANT PEB-DDEB CYS-2623.  
 RX MEDLINE-96081220; PubMed-8541842;  
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;  
 RT "Prebital epidermolysis bullosa: genetic linkage to COL7A1 and  
 RT identification of a glycine-to-cysteine substitution in the triple-  
 RT helical domain of type VII collagen.";  
 RL Hum. Mol. Genet. 4:1579-1583(1995).  
 RN [12]  
 RP VARIANT DDEB ARG-2043.  
 RX MEDLINE-95164985; PubMed-7861014;  
 RA Christiano A.M., Morriconi A., Paradiisi M., Angelo C., Mazzanti C.,  
 RA Cavalieri R., Uitto J.;  
 RT "A glycine-to-arginine substitution in the triple-helical domain of  
 RT type VII collagen in a family with dominant dystrophic epidermolysis  
 RT bullosa.";  
 RL J. Invest. Dermatol. 104:438-440(1995).  
 RN [13]  
 RP VARIANTS RDEB AND DDEB.  
 RX MEDLINE-96220218; PubMed-8644729;  
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;  
 RT "Glycine substitutions in the triple-helical region of type VII  
 RT collagen result in a spectrum of dystrophic epidermolysis bullosa  
 RT phenotypes and patterns of inheritance.";  
 RL Am. J. Hum. Genet. 58:671-681(1996).  
 RN [14]  
 RP VARIANT RDEB ARG-2575.  
 RX MEDLINE-96154068; PubMed-8592061;  
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;  
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:  
 RT genotype/phenotype correlation in a case of moderate clinical  
 RT severity.";  
 RL J. Invest. Dermatol. 106:119-124(1996).  
 RN [15]  
 RP VARIANT RDEB ARG-1782.  
 RX MEDLINE-96183562; PubMed-8618018;  
 RA Christiano A.M., McGrath J.A., Uitto J.;  
 RT "Influence of the second COL7A1 mutation in determining the  
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 106:766-770(1996).  
 RN [16]  
 RP VARIANT RDEB ASP-2073.  
 RX MEDLINE-96310789; PubMed-8757758;  
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,  
 RA Uitto J., Pope F.M., Eady R.A.J.;  
 RT "Clinicopathological correlations of compound heterozygous COL7A1  
 RT mutations in recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 107:171-177(1996).  
 RN [17]  
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 AND  
 RP R-2575.  
 RX MEDLINE-97465605; PubMed-9326325;  
 RA Hovanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,  
 RA Fralangi S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,  
 RA de Prost Y.;  
 RT "Characterization of 18 new mutations in COL7A1 in recessive  
 RT dystrophic epidermolysis bullosa provides evidence for distinct  
 RT molecular mechanisms underlying defective anchoring fibril  
 RT formation.";  
 RL Am. J. Hum. Genet. 61:599-610(1997).  
 RN [18]  
 RP VARIANT RDEB ARG-1652.  
 RX MEDLINE-98106792; PubMed-9444387;  
 RA Cserhalmi-Friedman P.B., Karpatt S., Horvath A., Christiano A.M.;  
 RT "Identification of a glycine substitution and a splice site mutation  
 RT in the type VII collagen gene in a proband with milds recessive  
 RT dystrophic epidermolysis bullosa.";  
 RL Arch. Dermatol. Res. 289:640-645(1997).  
 RN [19]  
 RP VARIANTS DEB ARG-2009 AND ARG-2043.  
 RX MEDLINE-97358588; PubMed-9215684;  
 RA Winberg J.-O., Hammami-Hausali N., Nilssen O., Anton-Lamprecht I.,  
 RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,

RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;  
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by  
 RT a splice site mutation in combination with a missense mutation in the  
 RT COL7A1 gene.";  
 RL Hum. Mol. Genet. 6:1125-1135(1997).  
 RN [20]  
 RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.  
 RX MEDLINE-98334662; PubMed-9668111;  
 RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luechi U.,  
 RA Luger T., Bruckner-Tuderman L.;  
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in  
 RT intracellular accumulation of collagen VII, loss of anchoring  
 RT fibrils, and skin blistering.";  
 RL J. Biol. Chem. 273:19228-19234(1998).  
 RN [21]  
 RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.  
 RX MEDLINE-98410969; PubMed-9740253;  
 RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Uitto J.;  
 RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis  
 RT bullosa.";  
 RL J. Invest. Dermatol. 111:534-537(1998).  
 RN [22]  
 RP VARIANT RDEB ARG-1347.  
 RX MEDLINE-99019477; PubMed-9804332;  
 RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,  
 RA Zamburano G., Bruckner-Tuderman L., Castiglia D.;  
 RT "Compound heterozygosity for a recessive glycine substitution and a  
 RT splice site mutation in the COL7A1 gene causes an unusually mild form  
 RT of localized recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 111:744-750(1998).  
 RN [23]  
 RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.  
 RX MEDLINE-99072663; PubMed-9856843;  
 QY 1 GVKDKNPFGPCAP 15  
 Db 2386 GVKDGLPLGLPGAP 2400  
 Query Match 68.5%; Score 63; DB 1; Length 2944;  
 Best local Similarity 80.0%; Pred. No. 0.2;  
 Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 RESULT 8  
 CA34\_HUMAN  
 ID CA34\_HUMAN STANDARD; PRT; 1670 AA.  
 AC 001955;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).  
 GN COL4A3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE-94364994; PubMed-8083201;  
 RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeder S.T.;  
 RT "Complete primary structure of the human alpha 3(IV) collagen chain.  
 RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  
 RT human tissues.";  
 RL J. Biol. Chem. 269:23013-23017(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Leinonen A.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE-93015826; PubMed-1400291;  
 RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;

RT RT "Exon/Intron structure of the human alpha 3(IV) gene encompassing the  
RT RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
RT RT antigenic region at the triple helix/NC1 domain junction.";  
RL RL J. Biol. Chem. 267:19780-19784(1992).  
RN RN [14]  
RN RP SEQUENCE OF 1453-1670 FROM N.A.  
RX RX MEDLINE=91353570; PubMed=1882840;  
RA RA Morrison K.E., Mariyama M., Yang-Feng T.L., Reeder S.T.;  
RT RT "Sequence and localization of a partial cDNA encoding the human alpha  
RT RT 3 chain of type IV collagen.";  
RL RL Am. J. Hum. Genet. 49:545-554(1991).  
RN RN [15]  
RN RP SEQUENCE OF 1331-1670 FROM N.A.  
RC RC TISSUE-Kidney;  
RX RX MEDLINE=9214/878; PubMed=1737849;  
RA RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
RA Pusey C.D.;  
RT RT "Molecular cloning of the human Goodpasture antigen demonstrates it  
RT RT to be the alpha 3 chain of type IV collagen.";  
RL RL J. Clin. Invest. 89:592-601(1992).  
RN RN [16]  
RN RP SEQUENCE OF 1644-1670 FROM N.A.  
RC RC TISSUE-Kidney;  
RA RA Ding J.;  
RL RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN RN [17]  
RN RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.  
RC RC TISSUE-Kidney;  
RX RX MEDLINE=94124597; PubMed=8294492;  
RA RA Feng L., Xia Y., Wilson C.B.;  
RT RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
RT RT collagen gene. Differential expression of mRNA transcripts that  
RT RT predict three protein variants with distinct carboxyl regions.";  
RL RL J. Biol. Chem. 269:2342-2348(1994).  
RN RN [18]  
RN RP SEQUENCE OF 1-29 FROM N.A.  
RX RX MEDLINE=98196854; PubMed=9537506;  
RA RA Momota R., Sugimoto M., Ohashi T., Kiyasawa K., Yoshioka H.,  
RA Minomaya Y.;  
RT RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
RT RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
RT RT 2q36.";  
RL RL FEBS Lett. 424:11-16(1998).  
RN RN [19]  
RN RP ALTERNATIVE SPLICING.  
RX RX MEDLINE=93280184; PubMed=8505332;  
RA RA Bernal D., Quinones S., Saus J.;  
RT RT "The human mRNA encoding the Goodpasture antigen is alternatively  
RT RT spliced.";  
RL RL J. Biol. Chem. 268:12090-12094(1993).  
RN RN [10]  
RN RP VARIANT PRO-1474.  
RX RX MEDLINE=95078827; PubMed=7987301;  
RA RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
RA Barrileto A., Monens L.A.H., van Oost B.A., Brunner H.G.,  
RA Reenders S.T., Smeets H.J.M.;  
RT RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
RT RT recessive Alport syndrome.";  
RL RL Hum. Mol. Genet. 3:1269-1273(1994).  
RN RN [11]  
RN RP VARIANTS AS E-297; R-407; R-640; R-1167; E-1207; Q-1215; S-1277; T-  
RN RP 1330; E-1334; E-1347 and C-1651, AND VARIANTS R-43; E-162; Y-326; H-  
RP RP 408; R-451; L-574; E-1269 AND P-1474.  
RX RX MEDLINE=21064696; PubMed=11134255;  
RA RA Heidel L., Arrondeel C., Forestier L., Cohen-Solal L., Mollet G.,  
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
RT RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
RT RT autosomal Alport syndrome.";  
RL RL J. Am. Soc. Nephrol. 12:907-106(2001).  
RN RN [-1-]  
RN RP FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
CC CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
CC CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
CC CC NIDOGEN

CC	-	SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC	-	ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC	-	WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC	-	SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC	-	ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/V AND
CC	-	3/4/5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC	-	C-TERMINAL NCI DOMAINS.
CC	-	TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC	-	COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC	-	COCHLEA, LUNG AND BRAIN.
CC	-	DOMAIN: NCPI AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC	-	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC	-	CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC	-	TRIPLE-HELICAL 7S DOMAIN.
CC	-	PWM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC	-	UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	-	PPM: THE ALTERNATIVE SPLICED FORN V CONTAINS AN ADDITIONAL
CC	-	N-LINKED GLYCOSYLATION SITE.
CC	-	PWM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC	-	ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC	-	THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC	-	IV COLLAGENS.
CC	-	DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
CC	-	THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CC	-	CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC	-	DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
CC	-	AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
CC	-	GLOMERULOEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC	-	HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC	-	MALES AND FEMALES.
CC	-	-----
CC	-	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	-	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	-	the European Bioinformatics Institute. There are no restrictions on its
CC	-	use by non-profit institutions as long as its content is in no way
CC	-	modified and this statement is not removed. Usage by and for commercial
CC	-	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/ or_send_an_email_to_license@isb-sib.ch</a> ).
CC	-	-----
DR	EMBL:	X80031: CAAS6335.1; -
DR	EMBL:	M92893: AAA21610.1; -
DR	EMBL:	S53790: AAB19637.1; -
DR	EMBL:	M81379: AAAS156.1; -
DR	EMBL:	L08650: AAAS2044.1; -
DR	EMBL:	U02519: AAA18942.1; -
DR	EMBL:	U02520: AAA18943.1; -
DR	EMBL:	AB008495; BAA25064.1; -
DR	MIM:	120070; -
DR	MIM:	203780; -
DR	MIM:	233450; -
DR	InterPro:	IPR001442; C4.
DR	InterPro:	IPR000087; Collagen.
DR	pfam:	PF01413; C4; 2.
DR	pfam:	PF01391; Collagen; 21.
DR	ProDom:	PD003923; C4; 2.
DR	SMART:	SMO0111; C4; 2.
KW	Extracellular matrix:	Connective tissue; Repeat; Hydroxylation;
KW	Glycoprotein:	Basement membrane; Collagen; Signal; Cell adhesion;
KW	Alternative splicing;	Polymorphism; Phosphorylation; Disease mutation;
KW	Alport syndrome.	
FT	SIGNAL	1 28
FT	CHAIN	29 1670
FT	DOMAIN	29 42
FT	DOMAIN	43 1438
FT	DOMAIN	1439 1670
FT	DOMAIN	1427 1444
FT	SITE	1426 1427
FT	SITE	791 793
FT	SITE	996 998
FT	CELL ATTACHMENT SITE (POTENTIAL).	
FT	CELL ATTACHMENT SITE (POTENTIAL).	
FT	CELL ATTACHMENT SITE (POTENTIAL).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	7S DOMAIN.	
FT	TRIPLE-HELICAL REGION.	
FT	NONHELICAL REGION (NC1) (GOODPASTURE	
FT	ANTIGEN) (BY SIMILARITY).	
FT	EPIPOE (RECOGNIZED BY GOODPASTURE	
FT	ANTIBODIES).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	7S DOMAIN.	
FT	TRIPLE-HELICAL REGION.	
FT	NONHELICAL REGION (NC1) (GOODPASTURE	
FT	ANTIGEN) (BY SIMILARITY).	
FT	EPIPOE (RECOGNIZED BY GOODPASTURE	
FT	ANTIBODIES).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	7S DOMAIN.	
FT	TRIPLE-HELICAL REGION.	
FT	NONHELICAL REGION (NC1) (GOODPASTURE	
FT	ANTIGEN) (BY SIMILARITY).	
FT	EPIPOE (RECOGNIZED BY GOODPASTURE	
FT	ANTIBODIES).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	7S DOMAIN.	
FT	TRIPLE-HELICAL REGION.	
FT	NONHELICAL REGION (NC1) (GOODPASTURE	
FT	ANTIGEN) (BY SIMILARITY).	
FT	EPIPOE (RECOGNIZED BY GOODPASTURE	
FT	ANTIBODIES).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	SITE (POTENTIAL).	
FT	7S DOMAIN.	
FT	TRIPLE-HELICAL REGION.	
FT	NONHELICAL REGION (NC1) (GOODPASTURE	
FT	ANTIGEN) (BY SIMILARITY).	
FT	EPIPOE (RECOGNIZED BY GOODPASTURE	
FT	ANTIBODIES).	
FT	CLEAVAGE (BY COLLAGENASE)	
FT	(BY SIMILARITY).	
FT	S	

```

FT SITE 1154 1156 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1306 1308 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1345 1347 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1432 1434 CELL ATTACHMENT SITE (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MOD_RES 1435 1435 PHOSPHORYLATION (BY SIMILARITY).
FT DISULFID 1437 1437 PHOSPHORYLATION (BY SIMILARITY).
FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
FT DISULFID 1503 1511 BY SIMILARITY.
FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
FT DISULFID 1616 1622 BY SIMILARITY.
FT VARSPLIC 1586 1670 FTSGSGTGOALASPGSCLEEFRRASPLECHGRGTCNYNS
  NYSFWLASINPERMRKPIPTVYKAGELEKTIISRCVCK
  KRH -> KATSDINSGWIKRNNKSGVHEETLKIKKTA
  ELVEFLIKRVTMEHAVI (IN ISOFORM 2).
  GTLGSCLQRTTPTFLFCNNDVCFNFRADYSYWLSTPAL
  MPNMARITGRALEPYISRCVCEGPAIAVAHSOTDIPP
  CPHGWSLAKGSPFIMFTSAGSGTGOALASPGSCLEEFRA
  SPLECHGRGTCNTYNSYSFWLASINPERMRKPIPTSVK
  AGELEKTIISRCVCKKRH -> DALEFKVLRSP (IN
  ISOFORM 3).
  G -> R.
  G -> R.
  /FTID-VAR_011202.
  G -> E.
  VARIANT 43 43
  VARIANT 162 162
  Query Match 67.4%; Score 62; DB 1; Length 1670;
  Best Local Similarity 66.7%; Pred. NO. 0.16;
  Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKNPGMGPAP 15
Db 1189 GAKGDRGAPGPG 1203

RESULT 9
COAS_BPPRD STANDARD; PRT: 339 AA.
AC P22536;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Minor capsid protein (protein p5).
GN V.
OS Bacteriophage PRD1.
OC Viruses; dsDNA viruses, no RNA stage: Tectiviridae; Tectivirus.
OX NCBI_TaxID=10658;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.
RA MEDLINE=90320115; Pubmed=2196741;
RA Bamford J.K.H., Bamford D.H.;
RT "Capsomer proteins of bacteriophage PRD1, a bacterial virus with a
  membrane.";
RL Virology 177:445-451(1990).
RN [2]
RP COLLAGENOUS REGION.
RA MEDLINE=90206069; Pubmed=2320123;
RA Bamford D.H., Bamford J.K.H.;
RT "Collagenous proteins multiply.";
RL Nature 344:497-497(1990).
CC -1- FUNCTION: THE SHORT COLLAGEN-LIKE REGION IS PROPOSED TO ACT AS A
  TRIMERIZATION SIGNAL LEADING TO THE FULFILLMENT OF THE SYMMETRY
  REQUIREMENT OF THE MINOR CAPSOMER.
CC -1- SUBUNIT: THE PRD1 VIRION IS COMPOSED OF AN ICOSAEDRAL PROTEIN
  COAT, AN INNER PROTEIN-LIPID MEMBRANE, AND A dsDNA GENOME WHICH
  IS LOCATED INSIDE THE LIPID VESICLE. THE PROTEIN COAT CONSISTS
  MAINLY OF PROTEIN P3. THE MINOR PROTEIN P5 IS ALSO CONSIDERED TO
  CONSTITUTE PART OF THE PROTEIN COAT.
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
  or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: M69077; AAA32460.1; -
DR EMBL: M55568; AAA32446.1; -
DR PIR: B46345; B46345.
KW Coat protein.
FT INIT_MET 0
FT DOMAIN 1 121 DOMAIN-1.
FT DOMAIN 122 140 COLLAGEN-LIKE.
FT DOMAIN 141 339 DOMAIN-2.
SQ SEQUENCE 339 AA; 34318 MW; 184BCF77D08944F3 CRC64;

Query Match 66.3%; Score 61; DB 1; Length 339;
Best Local Similarity 76.9%; Pred. NO. 0.046;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDKNPGMGP 13
Db 123 GIKGDKGDPGAP 135

RESULT 10
CAIF_HUMAN STANDARD; PRT: 1603 AA.
AC Q07092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVI) chain precursor.
GN COL16A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92335339; Pubmed=1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 418-1603 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93203161; Pubmed=1284248;
RA Yamaguchi N., Kimura S., McBride O.W., Horl H., Yamada Y.,
  Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
  chain, alpha 1(XVI), consisting of repetitive collagenous domains and
  cysteine-containing non-collagenous segments.";
RL J. Biochem. 112:856-863(1992).
CC -1- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
  THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
  AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
  COMPLEX NETWORK OF RETICULAR FIBERS, ALSO LOCATED TO A FIBROBLAST
  LAYER BENEATH THIS DENSE LAYER, EXISTS IN TISSUES IN ASSOCIATION
  WITH OTHER TYPES OF COLLAGEN.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
  GESTATION, AND DECREASE AT TERM.
CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
  TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
  DOMAINS (NC10 TO NC1).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
  UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
  INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).

CC -----  
 DR EMBL: M92642; AAA58427.1; -;  
 DR EMBL: S57132; AAB25797.1; -;  
 DR PIR: S23810; S23810.  
 DR MIM: 120326; -;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01391; Collagen; 16.  
 DR Pfam: PF02210; TSPN; 1.  
 DR SMART: SMO0210; TSPN; 1.  
 KM Extracellular matrix; Connective tissue; Collagen; Hydroxylation;  
 KM Repeat; Signal.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1603 COLLAGEN ALPHA 1(XVI) CHAIN.  
 FT DOMAIN 22 374 NONHELIICAL REGION 10 (NC10).  
 FT DOMAIN 375 505 TRIPLE-HELICAL REGION 9 (COL9)  
 FT WITH 3 IMPERFECTIONS.  
 FT DOMAIN 506 520 NONHELIICAL REGION 9 (NC9).  
 FT DOMAIN 521 554 TRIPLE-HELICAL REGION 8 (COL8)  
 FT WITH 1 IMPERFECTION.  
 FT DOMAIN 555 571 NONHELIICAL REGION 8 (NC8).  
 FT DOMAIN 572 630 TRIPLE-HELICAL REGION 7 (COL7)  
 FT WITH 1 IMPERFECTION.  
 FT DOMAIN 631 651 NONHELIICAL REGION 7 (NC7).  
 FT DOMAIN 652 722 TRIPLE-HELICAL REGION 6 (COL6)  
 FT WITH 1 IMPERFECTION.  
 FT DOMAIN 723 737 NONHELIICAL REGION 6 (NC6).  
 FT DOMAIN 738 875 TRIPLE-HELICAL REGION 5 (COL5)  
 FT WITH 3 IMPERFECTIONS.  
 FT DOMAIN 876 886 NONHELIICAL REGION 5 (NC5).  
 FT DOMAIN 887 938 TRIPLE-HELICAL REGION 4 (COL4)  
 FT WITH 2 IMPERFECTIONS.  
 FT DOMAIN 939 972 NONHELIICAL REGION 4 (NC4).  
 FT DOMAIN 973 987 TRIPLE-HELICAL REGION 3 (COL3).  
 FT DOMAIN 988 1010 NONHELIICAL REGION 3 (NC3).  
 FT DOMAIN 1011 1432 TRIPLE-HELICAL REGION 2 (COL2)  
 FT WITH 2 IMPERFECTIONS.  
 FT DOMAIN 1433 1471 NONHELIICAL REGION 2 (NC2).  
 FT DOMAIN 1472 1577 TRIPLE-HELICAL REGION 1 (COL1)  
 FT WITH 2 IMPERFECTIONS.  
 FT DOMAIN 1578 1603 NONHELIICAL REGION 1 (NC1).  
 FT DOMAIN 1603 418 RDA -> GGR (IN REF. 2).  
 FT CONFLICT 537 537 R -> P (IN REF. 2).  
 FT CONFLICT 1160 1160 T -> P (IN REF. 2).  
 FT CONFLICT 1163 1163 T -> P (IN REF. 2).  
 FT CONFLICT 1165 1165 S -> P (IN REF. 2).  
 SQ SEQUENCE 1603 AA; 157692 MW; E27D9A1D4E598A37 CRC64;

Query Match 66.3%; Score 61; DB 1; Length 1603;  
 Best Local Similarity 73.3%; Pred. No. 0.21;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPFGAP 15  
 DB 681 GOKGDAGNPGDPGP 695

RESULT 11  
 CAA4\_HUMAN STANDARD; PRT; 1690 AA.  
 AC P53420.  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 4(IV) chain precursor.  
 GN COL4A4.  
 OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISUE-Kidney;  
 RA MEDLINE=95014445; PubMed=7523402;  
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reedeers S.T.;  
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)  
 RT chain. Comparison with structure and expression of the other alpha  
 RT (IV) chains."; 269:26172-26177(1994).  
 RL J. Biol. Chem. 269:26172-26177(1994).  
 RN [2]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RA MEDLINE=98196854; PubMed=9537506;  
 RA Monota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 RT alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 RT 2q36.";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [3]  
 RP SEQUENCE OF 1219-1690 FROM N.A.  
 RC TISUE-Eye;  
 RA MEDLINE=93374047; PubMed=8365481;  
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;  
 RT "CDNA isolation and partial gene structure of the human alpha 4(IV)  
 RT collagen chain.";  
 RL FEBS Lett. 330:122-128(1993).  
 RN [4]  
 RP SEQUENCE OF 1407-1507 FROM N.A.  
 RA MEDLINE=93054733; PubMed=1429714;  
 RA Kamagata Y., Matel M.-G., Ninomiya Y.;  
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the  
 RT alpha 4 chain of basement membrane collagen type IV and assignment of  
 RT the gene to the distal long arm of human chromosome 2.";  
 RL J. Biol. Chem. 267:23753-23758(1992).  
 RN [5]  
 RP REVIEW ON VARIANTS.  
 RA MEDLINE=97338622; PubMed=9195222;  
 RA Lemlink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [6]  
 RP VARIANT AS SER-1201.  
 RA MEDLINE=95078927; PubMed=7987396;  
 RA Mochizuki T., Lemlink H.H., Mariyama M., Antignac C., Gubler M.-C.,  
 RA Pisoni Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,  
 RA Smeets H.J.M., Reedeers S.T.;  
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)  
 RT collagen genes in autosomal recessive Alport syndrome.";  
 RL Nat. Genet. 8:77-82(1994).  
 RN [7]  
 RP VARIANT BFN GLU-897.  
 RA MEDLINE=96379660; PubMed=8787673;  
 RA Lemlink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,  
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;  
 RT "Benign familial hematuria due to mutation of the type IV collagen  
 RT alpha4 gene.";  
 RL J Clin. Invest. 98:1114-1118(1996).  
 RN [8]  
 RP VARIANTS AS, AND VARIANTS.  
 RA MEDLINE=99011253; PubMed=9792860;  
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,  
 RA Cochot P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;  
 RT "Determination of the genomic structure of the COL4A4 gene and of  
 RT novel mutations causing autosomal recessive Alport syndrome.";  
 RL Am. J. Hum. Genet. 63:1329-1340(1998).  
 CC -I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCAN AND ENACTIN/  
 CC NIDOGEN.  
 CC -I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-

CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF TYPE II AUTOSOMAL  
 CC RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY  
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE A CAUSE OF FAMILIAL BENIGN  
 CC HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS  
 CC CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY  
 CC DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN  
 CC AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS  
 CC NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE  
 CC DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT  
 CC HEMATURIA AND THIN GBM AT THAT AGE.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X61053; CAA56943.1; -  
 CC EMBL; AB008496; BAA25065.1; -  
 CC EMBL; D17391; BAA04214.1; -  
 CC MIM; 120131; -  
 CC MIM; 141200; -  
 CC MIM; 203780; -  
 CC InterPro: IPR001442; C4.  
 CC InterPro: IPR000087; Collagen.  
 CC Pfam: PF01413; C4; 2.  
 CC Pfam: PF01391; Collagen; 20.  
 CC ProDom: PD003923; C4; 2.  
 CC SMART; SM00111; C4; 2.  
 CC Extracellular matrix; Connective tissue; Basement membrane; Repeat;  
 CC Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;  
 CC Polymorphism; Alport syndrome.  
 CC FT SIGNAL 1 38  
 CC FT CHAIN 39 1690  
 CC FT DOMAIN 39 64  
 CC FT DOMAIN 65 1459  
 CC FT DOMAIN 1460 1690  
 CC FT SITE 94 96  
 CC FT SITE 145 147  
 CC FT SITE 189 191  
 CC FT SITE 310 312  
 CC FT SITE 724 726  
 CC FT SITE 785 787  
 CC FT SITE 989 991  
 CC FT SITE 1206 1206  
 CC FT SITE 1212 1214  
 CC FT DISULFID 1480 1569  
 CC FT DISULFID 1513 1566  
 CC FT DISULFID 1525 1531

FT DISULFID 1588 1686  
 FT DISULFID 1622 1683  
 FT DISULFID 1634 1641  
 FT CARBOHYD 142 142  
 FT CARBOHYD 669 669  
 FT CARBOHYD 441 446  
 FT VARIANT 545 545  
 FT VARIANT 570 570  
 FT VARIANT 897 897  
 FT VARIANT 931 931  
 FT VARIANT 1004 1004  
 FT VARIANT 1030 1030  
 FT VARIANT 1201 1201  
 FT VARIANT 1402 1402  
 FT VARIANT 1572 1572  
 FT CONFLICT 1659 1660  
 SQ SEQUENCE 1690 AA; 164095 MR; EIE72P28A72BAAE CXC64;  
 Query Match 66.3%; Score 61; DB 1; Length 1690;  
 Best Local Similarity 60.0%; Pred. No. 0.22;  
 Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GVGDKGNPCWPCAP 15  
 DB 184 GIGQDRDPLGLPLP 198  
 RESULT 12  
 CA24\_ASCSU  
 ID CA24\_ASCSU STANDARD; PRT; 1763 AA.  
 AC P27393;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Collagen alpha 2(IV) chain precursor.  
 OS Ascaris suum (pig roundworm) (Ascaris lumbricoides).  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascarididae;  
 OC Ascarididae; Ascaris.  
 OX NCBI\_TaxId=6253;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-91340768; PubMed-1714907;  
 RA Pettit J., Kingston I.B.;  
 RT "The complete primary structure of a nematode alpha 2(IV) collagen  
 RT and the partial structural organization of its gene."  
 RL J. Biol. Chem. 266:16149-16156(1991).  
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.  
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.  
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH  
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1  
 CC DOMAINS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE







```

RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Nishimura Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT alpha 4 chain of basement membrane collagen type IV and assignment of
RT the gene to the distal long arm of human chromosome 2."
RL J. Biol. Chem. 267:23753-23758(1992).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLomerular basement membranes (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOCEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L01477; -; NOT_ANNOTATED_CDS.
CC DR Interpro: IPR001442; C4.
CC DR Interpro: IPR000087; Collagen.
CC DR Pfam: PF01413; C4; 2.
CC DR Pfam: PF01391; Collagen; 5.
CC DR ProDom: PD003923; C4; 2.
CC DR SMART: SM00111; C4; 2.
CC KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
CC FT NON_TER 1
CC FT DOMAIN <1 392 TRIPLE-HELICAL REGION.
CC FT PROPEP 393 623 NONHELICAL REGION (NC1).
CC FT DISULFID 413 502 OR 499 (BY SIMILARITY).
CC FT DISULFID 446 499 OR 502 (BY SIMILARITY).
CC FT DISULFID 458 464 BY SIMILARITY.
CC FT DISULFID 521 619 OR 616 (BY SIMILARITY).
CC FT DISULFID 555 616 OR 619 (BY SIMILARITY).
CC FT DISULFID 567 574 BY SIMILARITY.
CC SQ SEQUENCE 623 AA; 62393 MW; 62393B31242F882 CRC64;
Query Match 65.2%; Score 60; DB 1; Length 623;
Best Local Similarity 76.9%; Pred. No. 0.12;
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
OY 1 GYKDGKGNPGMPG 13
DB 139 GIKDGKGPSPG 151
RESULT 15
CA13_RAT
ID CA13_RAT STANDARD: PRT; 636 AA.
AC P13941; 070604;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain (fragment).
GN COL3A1.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94114571; PubMed=8286415;
RX Glumoff V., Maekela J.K., Vuorio E.;
RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different
RT expression patterns of type I and type III collagen and fibronectin
RT genes in experimental granuloma tissue."
RL Biochim. Biophys. Acta 1217:41-48(1994).
RN [2]
RP SEQUENCE OF 73-636 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Fibroblast;
RA Wirtz T., Ellerstrom C., Lundmark C., Christerson C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 308-482 FROM N.A.
RX MEDLINE=88296083; PubMed=2456904;
RA Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lytle C.R.,
RA Komm B., Mohr K.;
RT "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen
RT mRNAs by estradiol in the immature rat uterus."
RL DNA 7:347-354(1988).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X70369; CAA49832.1; -
CC DR EMBL: A0005395; CAA06510.1; -
CC DR EMBL: M2154; AAA40942.1; -
CC DR PIR: A29905; A29905.
CC DR PIR: S41067; S41067.
CC DR Interpro: IPR000087; Collagen.
CC DR Interpro: IPR000885; Fib_Collagen_C.
CC DR Interpro: IPR001007; WMFC.
CC DR Pfam: PF01410; COLF1; 1.
CC DR Pfam: PF01391; Collagen; 6.
CC DR ProDom: PD002078; Fib_Collagen_C; 1.
CC DR SMART: SM00038; COLF1; 1.
CC DR PROSITE: PS01208; WMFC; PARTIAL.
CC KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein.
CC FT NON_TER 1
CC FT CHAIN <1 375
CC FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.
CC FT DOMAIN <1 368 TRIPLE-HELICAL REGION.
CC FT DISULFID 369 636 NONHELICAL REGION (C-TERMINAL).
CC FT DISULFID 368 368 INTERCHAIN (BY SIMILARITY).
CC FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).
CC FT CONFLICT 340 340 N -> D (IN REF. 2).
CC FT CONFLICT 429 429 A -> G (IN REF. 2).
CC SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;
Query Match 65.2%; Score 60; DB 1; Length 636;
Best Local Similarity 71.4%; Pred. No. 0.12;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
OY 1 GYKDGKGNPGMPGA 14
IIII::III III

```

DB 30 GVKGERGSPGPGA 43

RESULT 16

CA13\_BOVIN STANDARD: PRT: 1049 AA.

ID CA13\_BOVIN

AC P04258;

DT 20-MAR-1987 (Rel. 04, Created)

DT 20-MAR-1987 (Rel. 04, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 1(III) chain.

CN COL3A1.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE OF 1-242.

RX MEDLINE=80026026; PubMed=488906;

RA Fietzek P.P., Allmann H., Rauterberg J., Henkel W., Wachter E., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. I. The amino acid sequence of the amino terminal region of the alpha 1(III) chain (positions 1-222)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).

RN [2]

RP SEQUENCE OF 243-422.

RX MEDLINE=80026027; PubMed=488907;

RA Dewes H., Fietzek P.P., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. II. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2 (positions 223-402)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).

RN [3]

RP SEQUENCE OF 423-571.

RX MEDLINE=80026028; PubMed=488908;

RA Bentz H., Fietzek P.P., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. III. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4 (positions 403-551)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).

RN [4]

RP SEQUENCE OF 572-808.

RX MEDLINE=80026029; PubMed=488909;

RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. IV. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB5 (positions 552-788)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).

RN [5]

RP SEQUENCE OF 809-947.

RX MEDLINE=80026030; PubMed=488910;

RA Dewes H., Fietzek P.P., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. V. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A (position 789-927)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).

RN [6]

RP SEQUENCE OF 948-1049.

RX MEDLINE=80026031; PubMed=488911;

RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;

RT "The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxyterminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028)."

RT Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).

RL -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN.

CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLISINES.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

DR PIR: A02862; CGB07S.

DR InterPro: IPR000087; Collagen.

DR InterPro: IPR001007; WMFC.

DR Pfam: PF01391; Collagen: 17.

DR PROSITE: PS01208; WMFC; PARTIAL.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen.

FT DOMAIN 1 14

FT DOMAIN 15 1040

FT DOMAIN 1041 1049

FT MOD\_RES 95 95

FT MOD\_RES 107 107

FT MOD\_RES 119 119

FT MOD\_RES 938 938

FT MOD\_RES 950 950

FT CARBOHYD 107 107

FT CARBOHYD 950 950

FT DISULFID 1040 1040

FT DISULFID 1041 1041

SQ SEQUENCE 1049 AA; 93651 MW; 8BEC33DIC66EC9A3 CRC64;

Query Match 65.2%; Score 60; DB 1; Length 1049;

Best Local Similarity 71.4%; Pred. No. 0.2;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGMPGA 14

DB 702 GVKGERGSPGPGA 715

RESULT 17

CA13\_CHICK STANDARD: PRT: 1262 AA.

ID CA13\_CHICK

AC P12105; P79758; P79759; Q90794; Q92029;

DT 01-OCT-1989 (Rel. 12, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Collagen alpha 1(III) chain precursor (Fragments).

GN COL3A1.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxID=9031;

RN [1]

RP SEQUENCE OF 1-886 FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=94265842; PubMed=8206952;

RA Nah H.-D., Niu Z., Adams S.L.;

RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen."

RT J. Biol. Chem. 269:16443-16448(1994).

RN [2]

RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.

RX MEDLINE=84270696; PubMed=6547770;

RA Yamada Y., Liao G., Mudryj M., Obici S., de Crombrughe B.;

RT "Conservation of the sizes for one but not another class of exons in two chick collagen genes."

RL Nature 310:333-337(1984).

RN [3]

RP SEQUENCE OF 977-1262 FROM N.A.

RX MEDLINE=83220816; PubMed=6856474;

RA Yamada Y., Kuhn K., de Crombrughe B.;

RT "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes."

RL Nucleic Acids Res. 11:2733-2744(1983).

CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN.

CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLISINES.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING

CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 1 WMFC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U07973; AAA83407.1; -  
 CC EMBL: X00822; CAB52686.1; -  
 CC EMBL: X00823; CAB52686.1; JOINED.  
 CC EMBL: X00826; CA253397.1; ALT\_SEQ.  
 CC EMBL: X00825; CA253397.1; JOINED.  
 CC EMBL: X00827; CA253396.1; -  
 CC EMBL: X00828; CA253399.1; -  
 CC EMBL: X00830; CA25401.1; -  
 CC EMBL: X00831; CA25402.1; -  
 CC EMBL: K02301; AAD15298.1; -  
 CC EMBL: M36662; AAA18519.1; ALT\_SEQ.  
 CC PIR: A05269; A05269.  
 CC InterPro: IPR000087; Collagen.  
 CC InterPro: IPR000885; Fib.collagen\_C.  
 CC InterPro: IPR001007; WMFC.  
 CC Pfam: PF01381; Collagen; 13.  
 CC ProDom: PD002078; Fib.collagen\_C; 1.  
 CC SMART: SM00038; COLFI; 1.  
 CC SMART: SM00214; WVC; 1.  
 CC PROSITE: PS01208; WVC; 1.  
 CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 CC Glycoprotein; Collagen; Signal.  
 CC FT SIGNAL 1 23  
 CC FT PROPEP 24 144  
 CC FT CHAIN 145 1003  
 CC FT PROPEP 1004 1262  
 CC FT DOMAIN 29 88  
 CC FT DOMAIN 145 164  
 CC FT DOMAIN 165 994  
 CC FT DOMAIN 995 1003  
 CC FT NON\_CONS 886 887  
 CC FT NON\_CONS 922 923  
 CC FT DISULFID 994 994  
 CC FT DISULFID 995 995  
 CC FT MOD\_RES 262 262  
 CC FT MOD\_RES 283 283  
 CC FT MOD\_RES 859 859  
 CC FT CARBOHYD 1163 1163  
 CC FT CONFLICT 96 96  
 CC FT CONFLICT 1132 1132  
 CC SO SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DBA43D CMC64;  
 CC  
 CC Query Match 65.2%; Score 60; DB 1; Length 1262;  
 CC Best Local Similarity 71.4%; Pred. No. 0.24;  
 CC Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain precursor.  
 GN COL3A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Skin fibroblast;  
 RX MEDLINE-89350838; PubMed-276486;  
 RA Ala-Koko L., Kontusaari S., Baldwin C.T., Kulvanenti H.,  
 RA Prockop D.J.;  
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
 RT chain of human type III procollagen. Differences in protein structure  
 RT from type I procollagen and conservation of codon preferences.";  
 RL Biochem. J. 260:509-516(1989).  
 RN [2]  
 RP SEQUENCE OF 149-1225 FROM N.A.  
 RX MEDLINE-89386015; PubMed-2780304;  
 RA Janeczko R.A., Ramirez F.;  
 RT "Nucleotide and amino acid sequences of the entire human alpha 1  
 RT (III) collagen.";  
 RL Nucleic Acids Res. 17:6742-6742(1989).  
 RN [3]  
 RP SEQUENCE OF 168-398.  
 RX MEDLINE-77134724; PubMed-557335;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of cyanogen  
 RT bromide peptides from the amino-terminal segment of type III collagen  
 RT of human liver.";  
 RL Biochemistry 16:1158-1164(1977).  
 RN [4]  
 RP REVISIONS.  
 RA Seyer J.M.;  
 RL Submitted (DEC-1977) to the PIR data bank.  
 RN [5]  
 RP SEQUENCE OF 399-727.  
 RX MEDLINE-79000343; PubMed-687591;  
 RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of five  
 RT consecutive CNBr peptides from type III collagen of human liver.";  
 RL Biochemistry 17:3404-3411(1978).  
 RN [6]  
 RP SEQUENCE OF 728-964.  
 RX MEDLINE-80198282; PubMed-6246925;  
 RA Seyer J.M., Mainardi C., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha 1  
 RT (III)-CB5 from type III collagen of human liver.";  
 RL Biochemistry 19:1583-1589(1980).  
 RN [7]  
 RP SEQUENCE OF 950-1466 FROM N.A.  
 RX MEDLINE-88189827; PubMed-3357782;  
 RA Mankoo B.S., Dalgleish R.;  
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RL Nucleic Acids Res. 16:2337-2337(1988).  
 RN [8]  
 RP REVISION TO 1184.  
 RX MEDLINE-89098346; PubMed-3211760;  
 RA Molyneux K., Dalgleish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:11833-11833(1988).  
 RN [9]  
 RP SEQUENCE OF 1065-1466 FROM N.A.  
 RX MEDLINE-85087944; PubMed-6096827;  
 RA Ioldi H.R., Brinker J.M., May M., Phlajantemi T., Morrow S.,  
 RA Rosenbloom J., Myers J.C.;  
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
 RT procollagen.";  
 RL Nucleic Acids Res. 12:9383-9394(1984).  
 RN [10]  
 RP SEQUENCE OF 965-1200.  
 RX MEDLINE-81208139; PubMed-7016180;

RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha  
 RT 1(III)-CB9 from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 [11]  
 RP SEQUENCE OF 1176-1466 FROM N.A.  
 RA MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippoia M., Ramirez F.;  
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1  
 RT (III) collagen. Partial characterization of the 3' end region of the  
 RT gene.";  
 RL J. Biol. Chem. 260:4357-4363(1985).  
 [12]  
 RP SEQUENCE OF 1161-1200 FROM N.A.  
 RA MEDLINE=86187804; PubMed=3754462;  
 RA Miskulin M., Dalgleish R., Kluwe-Beckerman B., Rennard S.I.,  
 RA Tolstoshev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated  
 RT with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 [13]  
 RP SEQUENCE OF 1-170 FROM N.A.  
 RC TISSUE-Placenta;  
 RA MEDLINE=88303360; PubMed=3405773;  
 RA Toman D., Ricca G., de Crombrughe B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
 RT of human proalpha 1(III) collagen.";  
 RL Nucleic Acids Res. 16:7201-7201(1988).  
 [14]  
 RP SEQUENCE OF 1-176 FROM N.A.  
 RA MEDLINE=89378752; PubMed=2777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III  
 RT procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1989).  
 [15]  
 RP REVIEW ON VARIANTS.  
 RA MEDLINE=97255959; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 [16]  
 RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RA MEDLINE=93293988; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhathri S.L., Kleihert C.,  
 RA Earley J.J., Zhang J., Noerregaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Rynnaenen M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smulders S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
 RT in the triple-helical domain of type III procollagen are an  
 RT infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 [17]  
 RP VARIANT THR-698.  
 RA MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Kleihert C., Tromp G., Kuivaniemi H., Kontusaari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 [18]  
 RP VARIANT AORTIC ANEURYSM ARG-786.  
 RA MEDLINE=91056145; PubMed=2243125;  
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
 RT with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 [19]  
 RP VARIANT EDS-IV ARG-828.  
 RA MEDLINE=94016385; PubMed=8411057;  
 RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;

RT "The substitution of glycine 661 by arginine in type III collagen  
 RT produces mutant molecules with different thermal stabilities and  
 RT causes Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 30:690-693(1993).  
 [20]  
 RP VARIANT EDS-IV SER-957.  
 RA MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of  
 RT the alpha 1 (III) chain of type III procollagen exposes an arginine  
 RT and causes Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:1349-1352(1989).  
 [21]  
 RP VARIANT EDS-IV VAL-960.  
 RA MEDLINE=95268429; PubMed=7749417;  
 RA Tromp G., de Paape A., Nuytink L., Madhathri S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in  
 RT Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 [22]  
 RP VARIANT EDS-IV GLU-1014.  
 RA MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
 RA Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1)  
 RT converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
 RT syndrome type IV. An unaffected family member is mosaic for the  
 RT mutation.";  
 RL Hum. Genet. 89:414-418(1992).  
 [23]  
 RP VARIANT EDS-IV ASP-1050.  
 RA MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts  
 RT the codon for glycine 883 to aspartate in a mild variant of  
 RT Ehlers-Danlos syndrome IV.";  
 RL J. Biol. Chem. 264:19313-19317(1989).  
 [24]  
 RP VARIANT EDS-IV VAL-1077.  
 RA MEDLINE=91374480; PubMed=1895316;  
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paape A., Narcisi P.,  
 RA Pope F.M.;  
 RT "Characterisation of a glycine to valine substitution at amino acid  
 RT position 910 of the triple helical region of type III collagen in a  
 RT patient with Ehlers-Danlos syndrome type IV.";  
 RL J. Med. Genet. 28:458-463(1991).  
 [25]  
 RP VARIANT EDS-IV GLU-1173.  
 RA MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
 RT Query Match 65.2%; Score 60; DB 1; Length 1466;  
 RT Best Local Similarity 78.6%; Pred. No. 0.28; Mismatches 3; Indels 0;  
 RT Matches 11; Conservative 0; Gaps 0;  
 QY 1 GVKGDKNPMPGA 14  
 | | | | | | | | | |  
 DB 738 GPKGDKEGPGPGA 751  
 RESULT 19  
 ID CA34-BOVIN STANDARD; PRT; 471 AA.  
 AC 028084;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Collagen alpha 3(IV) chain (Fragment).  
 GN COL4A3.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lens;  
 RX MEDLINE=91093146; PubMed=1985905;  
 RA Morrison K.E., Gerlino G.G., Reiders S.T.;  
 RT "Use of the polymerase chain reaction to clone and sequence a cDNA  
 RL J. Biol. Chem. 266:34-39(1991).  
 RN [2]  
 RP SEQUENCE OF 227-258.  
 RC TISSUE=Kidney;  
 RX MEDLINE=90202779; PubMed=2318822;  
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;  
 RT "Glomerular basement membrane. Identification of a fourth chain,  
 RL J. Biol. Chem. 265:5466-5469(1990).  
 RN [3]  
 RP SEQUENCE OF 227-254.  
 RX MEDLINE=88330844; PubMed=3417661;  
 RA Saus J., Wieselander J., Langeveld J.P.M., Quinones S., Hudson B.G.;  
 RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain  
 RL J. Biol. Chem. 263:13374-13380(1988).  
 RN [4]  
 RP SEQUENCE OF 227-244.  
 RX MEDLINE=87222419; PubMed=2438283;  
 RA Butkowski R.J., Langeveld J.P.M., Wieselander J., Hamilton J.,  
 RT "Localization of the Goodpasture epitope to a novel chain of basement  
 RT membrane collagen.";  
 RL J. Biol. Chem. 262:7874-7877(1987).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ACTIN/  
 CC NIDDOGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M63139; AAA62708.1; -  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 4.  
 DR Prodom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 FT Glycoprotein; Basement membrane; Collagen; Cell adhesion.  
 FT NON\_TER 1 238  
 FT DOMAIN <1 238 TRIPLE-HELICAL REGION.  
 FT DOMAIN 239 471 NONHELICAL REGION (NC1).

FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).  
 FT MOD\_RES 232 232 HYDROXYLATION.  
 FT MOD\_RES 238 238 HYDROXYLATION.  
 FT DISULFID 261 352 OR 349 (BY SIMILARITY).  
 FT DISULFID 294 349 OR 352 (BY SIMILARITY).  
 FT DISULFID 306 312 BY SIMILARITY.  
 FT DISULFID 371 466 OR 463 (BY SIMILARITY).  
 FT DISULFID 405 463 OR 466 (BY SIMILARITY).  
 FT DISULFID 417 423 BY SIMILARITY.  
 FT CONFLICT 253 253 S -> Y (IN REF. 3).  
 SQ SEQUENCE 471 AA: 47585 MW: C03B66F14E7008DE CRC64;  
 Query Match 64.1%; Score 59; DB 1; Length 471;  
 Best Local Similarity 66.7%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GYKDKGNPGWPGAP 15  
 DB 116 GKKKKGNSGFGPP 130  
 ID CA14\_CAEEL STANDARD; PRT; 1758 AA.  
 AC P17139;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN EMB-9 OR CLB-2 OR K0444.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=91141582; PubMed=1996137;  
 RA Guo X., Johnson J.J., Kramer J.M.;  
 RT "Embryonic lethality caused by mutations in basement membrane  
 RT collagen of C. elegans.";  
 RL Nature 349:707-709(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Almscough R., Anderson K., Baynes C., Barks M.,  
 RA Bonfield J., Burton J., Connell M., Copesey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rikken A., Roopra A., Saunders D., Showkeen R.,  
 RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Steden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,  
 RA Wohlfman P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 RN [3]  
 RP SEQUENCE OF 1446-1758 FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE=90008929; PubMed=2793871;  
 RA Guo X., Kramer J.M.;  
 RT "The two Caenorhabditis elegans basement membrane (type IV) collagen  
 RT genes are located on separate chromosomes.";  
 RL J. Biol. Chem. 264:17574-17582(1989).  
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.  
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.  
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH  
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1  
 CC DOMAINS.

```

CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- DISEASE: MUTATIONS IN THIS GENE CAUSE TEMPERATURE-SENSITIVE
CC LETHALITY DURING LATE EMBRYOGENESIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56979; CAA40299.1; -
DR EMBL: Z27078; CAA81584.1; -
DR EMBL: J05087; AAB59179.1; -
DR PIR: B34476; B34476.
DR PIR: S13651; S13651.
DR WormPep: K04H4.1; CE00246.
DR InterPro: IPR001442; C4.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 22.
DR ProDom: PD003923; C4; 2.
DR SMART: SM00111; C4; 2.
DR Extracellular matrix; Connective tissue; Basement membrane;
DR Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
DR SIGNAL: 1 20
DR PROPEP: 21 7194
DR CHAIN: 7195 1758 COLLAGEN ALPHA 1(IV) CHAIN.
DR DOMAIN: 1530 1529 NONHELICAL REGION (NC1).
DR DISULFID: 1530 1758 TRIPLE-HELICAL REGION (NC1).
DR DISULFID: 1582 1640 OR 1637 (BY SIMILARITY).
DR DISULFID: 1594 1600 OR 1640 (BY SIMILARITY).
DR DISULFID: 1659 1754 OR 1751 (BY SIMILARITY).
DR DISULFID: 1693 1751 OR 1754 (BY SIMILARITY).
DR DISULFID: 1705 1711 BY SIMILARITY.
DR VARIANT: 402 402 G -> E (IN MUTANT G23/HCT10).
DR VARIANT: 408 408 G -> E (IN MUTANT G23/HCT10).
DR VARIANT: 130 185 GPGMPGLAGPGQSGQNGNPRGLSGPPEGVNSQGR
KV GPGMPGLAGPGQSGQNGNPRGLSGPPEGVNSQGR
KV Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL: 1 20
FT PROPEP: 21 7194
FT CHAIN: 7195 1758 COLLAGEN ALPHA 1(IV) CHAIN.
FT DOMAIN: 1530 1529 NONHELICAL REGION (NC1).
FT DISULFID: 1530 1758 TRIPLE-HELICAL REGION (NC1).
FT DISULFID: 1582 1640 OR 1637 (BY SIMILARITY).
FT DISULFID: 1594 1600 OR 1640 (BY SIMILARITY).
FT DISULFID: 1659 1754 OR 1751 (BY SIMILARITY).
FT DISULFID: 1693 1751 OR 1754 (BY SIMILARITY).
FT DISULFID: 1705 1711 BY SIMILARITY.
FT VARIANT: 402 402 G -> E (IN MUTANT G23/HCT10).
FT VARIANT: 408 408 G -> E (IN MUTANT G23/HCT10).
FT VARIANT: 130 185 GPGMPGLAGPGQSGQNGNPRGLSGPPEGVNSQGR
FT GPGMPGLAGPGQSGQNGNPRGLSGPPEGVNSQGR
FT VLRVSLKCPDLIDLGNDKTETLDDSDSPHQEKVESIH
FT KDAKELKNELEDOEVOFO (IN REF. 2).
FT PGCPA -> LSDQV (IN REF. 2).
FT LDN -> AGQR (IN REF. 2).
FT MISSING (IN REF. 2).
FT G -> R (IN REF. 2).
FT P -> R (IN REF. 2).
FT PG -> TR (IN REF. 2).
FT D -> V (IN REF. 2).
FT P -> T (IN REF. 2).
FT P -> Q (IN REF. 3).
FT P -> L (IN REF. 2).
FT CONFLICT: 1514 1514 P -> L (IN REF. 2).
FT CONFLICT: 1722 1722 P -> L (IN REF. 2).
FT CONFLICT: 1752 1752 P -> L (IN REF. 2).
SQ SEQUENCE 1758 AA; 170857 MW; 7083D9AF63B05D45 CRC64;

```

Query Match Score 59; DB 1; Length 1758; Best Local Similarity 73.3%; Pred. No. 0.47; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GYKDGKNGPWPMPGAP 15  
DB 224 GYKDGKNGPWPMPGAP 238

```

RESULT 21
YRF2_CAEEL
AC YRF2_CAEEL STANDARD: PRT: 285 AA.
ID YRF2_CAEEL
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Putative cuticle collagen K03H9.2.
GN K03H9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Anderson K.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U21318; AAC46668.1; -
DR WormPep: K03H9.2; CE01996.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 2.
DR Hypothetical protein; Cuticle; Connective tissue; Repeat;
KV Collagen; Multigene family.
FT DOMAIN: 87 116 TRIPLE-HELICAL REGION.
FT DOMAIN: 133 261 TRIPLE-HELICAL REGION.
SQ SEQUENCE 285 AA; 30107 MW; F4619D41935A83FD CRC64;

```

Query Match Score 58; DB 1; Length 285; Best Local Similarity 60.0%; Pred. No. 0.11; Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GYKDGKNGPWPMPGAP 15  
DB 142 GIDGPGIIPGWPMPGAP 156

RESULT 22  
CA1H\_HUMAN  
ID CA1H\_HUMAN STANDARD: PRT: 1516 AA.  
AC P39060; Q9Y608; Q9Y607; Q9UK38;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 1(XVII) chain precursor [Contains: Endostatin].  
GN COL18A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98164096; PubMed=9503365;  
RA Saarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;  
RT "Complete primary structure of two variant forms of human type XVIII

RT collagen and tissue-specific differences in the expression of the  
RT corresponding transcripts".  
RL *Matrix Biol.* 16:319-328(1998).  
[2]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=402869759; PubMed=10830953;  
RX Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Onhi M., Takagi T., Sakaki Y., Tauden S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichenwald K., Rump A.J., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudoh Y., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordstiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario E., Reichelt J., Kauer G., Bloecher H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Resselmann L., Daggand E.,  
RA Reinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehrach H., Reinhardt R., Vaspou M.-L.,  
RA "The DNA sequence of human chromosome 21.";  
RL *Nature* 405:311-319(2000).  
[3]  
RN SEQUENCE OF 834-1516 FROM N.A.  
RP MEDLINE=94245237; PubMed=8188291;  
RX Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,  
RA Tissue-Placenta:  
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.,  
RT "Cloning and expression of human endostatin gene in *Escherichia*  
RT coli".  
RL submitted (SEP-1999) to the EMBL/GenBank/DBD databases.  
[5]  
RN INVOLVEMENT IN KNOBLOCH SYNDROME.  
RP MEDLINE=20400145; PubMed=10942434;  
RX Sertie A.L., Soesi V., Camargo A.A., Zatz M., Brahe C.,  
RA Passos-Bueno M.R.;  
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis  
RT and tumor growth, plays a critical role in the maintenance of retinal  
RT structure and in neural tube closure".  
RL *Hum. Mol. Genet.* 9:2051-2058(2000).  
[6]  
RN VARIANT ASN-1437.  
RP PubMed=11606364.  
RX Inghetti P., Suzuki O., Godol P.H., Alves V.A., Sertie A.L.,  
RA Zozick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,  
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;  
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes  
RT for the development of prostatic adenocarcinoma".  
RL *Cancer Res.* 61:7375-7378(2001).  
[7]  
RN FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE  
RP RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.  
RX FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
CC FACTOR SIGNALING.  
[8]  
RN ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM/NC1-303 AND A LONG  
RP FORM/NC-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
RX TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS  
CC IN LIVER, LUNG AND KIDNEY.  
[9]  
RN PPM: PROLINS AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
RP UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
RX POLYMORPHISM: There is an association between a polymorphism in  
CC position 1437 and prostate cancer. Heterozygous Asn-1437  
CC individuals have a 2.5 times increased chance of developing  
CC prostate cancer as compared with homozygous Asp-1437 individuals.  
CC [10]  
RN DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS  
RP or KNO) an autosomal recessive disorder defined by the occurrence  
CC of high myopia, vitreoretinal degeneration with retinal

CC		detrachment; mucular abnormalities and occipital encephalocle.
CC	-I-	SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.
CC		-----
CC		This SWISS-PROT entry is copyright. It is produced through a collabora-
CC		tion between the Swiss Institute of Bioinformatics and the EMBL outstation
CC		the European Bioinformatics Institute. There are no restrictions on
CC		use by non-profit institutions as long as its content is in no
CC		modified and this statement is not removed. Usage by and for commer-
CC		cials requires a license agreement (See <a href="http://www.isdb.ch/annou">http://www.isdb.ch/annou</a>
CC		or send an email to <a href="mailto:license@isdb-sib.ch">license@isdb-sib.ch</a> ).
CC		-----
DR	EMBL:	AF018082; AAC39658.1; -;
DR	EMBL:	AF018082; AAC39659.1; -;
DR	EMBL:	AL163302; CAB90482.1; -;
DR	EMBL:	L22548; AAA51864.1; -;
DR	EMBL:	AF184060; AAF01310.1; ALT_INIT.
DR	HSSP:	P39061; 1KOE.
DR	GlycoSuiteDB:	P39060; -.
DR	MIM:	120328; -.
DR	MIM:	267750; -.
DR	InterPro:	IPRO00087; Collagen.
DR	InterPro:	IPRO01791; LamlnInG.
DR	InterPro:	IPRO03129; TSPN.
DR	Pfam:	PF01391; Collagen; 6.
DR	Pfam:	PF02210; TSPN; 1.
DR	SMART:	SMO0282; Lamg; 1.
DR	SMART:	SMO0210; TSPN; 1.
KW	Extracellular matrix;	Connective tissue; Repeat; Hydroxylation;
KW	Cell adhesion;	Collagen; Glycoprotein; Signal; Alternative splicing;
KW	Polymorphism.	
FT	SIGNAL	1 23 POTENTIAL.
FT	CHAIN	24 1516 COLLAGEN ALPHA 1(XVIII) CHAIN.
FT	CHAIN	1334 1516 ENDOSTATIN.
FT	DOMAIN	24 516 NONHELTICAL REGION 1 (NC1).
FT	DOMAIN	517 550 TRIPLE-HELICAL REGION 1 (COL1).
FT	DOMAIN	551 560 NONHELTICAL REGION 2 (NC2).
FT	DOMAIN	561 640 TRIPLE-HELICAL REGION 2 (COL2).
FT	DOMAIN	641 664 NONHELTICAL REGION 3 (NC3).
FT	DOMAIN	665 786 TRIPLE-HELICAL REGION 3 (COL3).
FT	DOMAIN	787 809 NONHELTICAL REGION 4 (NC4).
FT	DOMAIN	810 892 TRIPLE-HELICAL REGION 4 (COL4).
FT	DOMAIN	893 906 NONHELTICAL REGION 5 (NC5).
FT	DOMAIN	907 948 TRIPLE-HELICAL REGION 5 (COL5).
FT	DOMAIN	949 961 NONHELTICAL REGION 6 (NC6).
FT	DOMAIN	962 1034 TRIPLE-HELICAL REGION 6 (COL6).
FT	DOMAIN	1035 1044 NONHELTICAL REGION 7 (NC7).
FT	DOMAIN	1045 1077 TRIPLE-HELICAL REGION 7 (COL7).
FT	DOMAIN	1078 1089 NONHELTICAL REGION 8 (NC8).
FT	DOMAIN	1090 1111 TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	1112 1118 NONHELTICAL REGION 9 (NC9).
FT	DOMAIN	1119 1173 TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	1174 1187 NONHELTICAL REGION 10 (NC10).
FT	DOMAIN	1187 1204 TRIPLE-HELICAL REGION 10 (COL10).
FT	DOMAIN	1205 1516 NONHELTICAL REGION 11 (NC11).
FT	CARBOHYD	68 68 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	129 129 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	164 164 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	691 691 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	CARBOHYD	1329 O-LINKED (GLAUGC. . . ) (POTENTIAL).
FT	CARBOHYD	/FTID-CAR_000150.
FT		BY SIMILARITY.
FT		D -> N.
FT	DISULFID	1366 1506 /FTID-VAR_012709.
FT	DISULFID	1468 1498 F -> S (IN REF. 2).
FT	SITE	1095 1097 I -> V (IN REF. 2).
FT	VARSPLIC	1 180 V -> L (IN REF. 3).
FT	VARSPLIC	181 215
FT		
FT	VARIANT	1437 1437
FT	CONFLICT	428 428
FT	CONFLICT	841 841
FT	CONFLICT	877 877



FT	CONFLICT	886	886	P -> R (IN REF. 3).
FT	CONFLICT	912	912	P -> R (IN REF. 3).
FT	CONFLICT	933	933	R -> L (IN REF. 3).
FT	CONFLICT	975	975	P -> L (IN REF. 3).
FT	CONFLICT	1064	1064	A -> P (IN REF. 3).
FT	CONFLICT	1084	1084	L -> K (IN REF. 3).
FT	CONFLICT	1120	1120	P -> A (IN REF. 3).
FT	CONFLICT	1123	1123	P -> A (IN REF. 3).
FT	CONFLICT	1126	1126	P -> PGCP (IN REF. 2).
FT	CONFLICT	1206	1206	G -> GQ (IN REF. 3).
FT	CONFLICT	1304	1304	R -> G (IN REF. 3).
FT	CONFLICT	1314	1314	A -> G (IN REF. 3).
FT	CONFLICT	1323	1324	LR -> CG (IN REF. 3).
FT	CONFLICT	1443	1443	R -> T (IN REF. 4).
FT	CONFLICT	1483	1483	S -> Y (IN REF. 4).
SQ	SEQUENCE	1516 AA:	153840 MM;	3C70FE29AA476EE76 CRC64;
QY		1 GVKGKGNPGMGAP 15		
		:		
Db		517 GLKGKGGEPGVPGR 531		
RESULT_23				
CA24_CAEEL				
ID CA24_CAEEL	STANDARD;	PRT,	1758 AA.	
AC p17140:				
DT 01-AUG-1990 (Rel. 15, Created)				
DT 01-OCT-1994 (Rel. 30, Last sequence update)				
DT 30-MAY-2000 (Rel. 39, Last annotation update)				
DE Collagen alpha 2(IV) chain precursor.				
LN LEF-2 OR CLB-1.				
OS Caenorhabditis elegans.				
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;				
OC Rhabdilitidae; Peloderinae; Caenorhabditis.				
OX NCBI_TaxID=6239;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN-BRISTOL N2;				
RA MEDLINE=94012964; Pubmed=7691828;				
RA Sidley M.H., Johnson J.J., Mello C.C., Kramer J.M.;				
RT "Genetic identification, sequence, and alternative splicing of the				
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";				
RL J. Cell Biol. 123:255-264(1993).				
RN [2]				
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.				
RC STRAIN-BRISTOL N2;				
RA MEDLINE=90089929; Pubmed=2793871;				
RA Guo X., Kramer J.M.;				
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen				
RT genes are located on separate chromosomes.";				
RL J. Biol. Chem. 264:17574-17582(1989).				
RN [3]				
RP VARIANTS.				
RA MEDLINE=94320591; Pubmed=8045258.				
RA Sidley M.H., Graham P.L., von Mendel N., Kramer J.M.;				
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of				
RT Caenorhabditis elegans produce phenotypes of differing severities.";				
RL EMO J. 13:3278-3285(1994).				
CC -I- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.				
CC -I- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.				
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH				
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1				
CC DOMAINS.				
CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE				
CC PRODUCED BY ALTERNATIVE SPLICING.				
CC -I- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM				
CC II IS PREDOMINANT IN THE LARVAE AND ADULTS.				
CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS				
CC DOMAIN (NC1) AT THEIR C-TERMINUS; FREQUENT INTERRUPTIONS OF THE				
CC				

CC	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC	TRIPLE-HELICAL 7S DOMAIN
CC	-I- PMM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	-I- PMM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRA-MOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
CC	-I- DISEASE: MUTATIONS IN LET-2 ARE GENERALLY EMBRYONIC LETHAL.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a> , or send an email to <a href="mailto:license@sdb.sdb.ch">license@sdb.sdb.ch</a> ).
CC	-----
DR	EMBL; Z22964; CAA80536.1; -;
DR	EMBL; Z22964; CAA80537.1; -;
DR	EMBL; J05066; AAA27989.1; -;
DR	EMBL; U22327; AAA64312.1; ALT_SEQ.
DR	PIR; A34476; A34476.
DR	InterPro; IPR001442; C4.
DR	InterPro; IPR000087; Collagen.
DR	Pfam; PF01413; C4; 2.
DR	Pfam; PF01391; Collagen; 23.
DR	ProDom; PD003923; C4; 2.
DR	SMART; SMO0111; C4; 2.
KW	Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen; Alternative splicing; Glycoprotein; Signal.
FT	SIGNAL 1 26
FT	CHAIN 27 1758
FT	DOMAIN 27 42
FT	DOMAIN 42 1527
FT	DOMAIN 1528 1758
FT	DISULFID 1546 1635
FT	DISULFID 1579 1632
FT	DISULFID 1591 1597
FT	DISULFID 1654 1750
FT	DISULFID 1688 1747
FT	DISULFID 1700 1707
FT	CARBOHYD 248 248
FT	VARSPLIC 229 264
FT	-----
FT	VARIANT 48 48
FT	VARIANT 366 366
FT	VARIANT 570 570
FT	VARIANT 588 588
FT	VARIANT 597 597
FT	VARIANT 690 690
FT	VARIANT 737 737
FT	VARIANT 877 877
FT	VARIANT 904 904
FT	VARIANT 1003 1003
FT	VARIANT 1125 1125
FT	VARIANT 1152 1152
FT	VARIANT 1286 1286
FT	CONFLICT 1682 1682
SO	SEQUENCE 1758 AA; 167750 MW; 97EB3FDBB2D2AC5 CRC64;
OY	Query Match 63.0%; Score 58; DB 1; Length 1758;
db	Best Local Similarity 73.3%; Pred. No. 0.66;
	Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
	1 GVKGDKGNPGRGAP 15
	416 GPKGDKGGGIGAP 430





```

RT "Complete primary structure of rainbow trout type I collagen
RT consisting of alpha1(I)alpha2(I)alpha3(I) heterotrimers."
RL Eur. J. Biochem. 268:2817-2827(2001).
RN [2]
RP SEQUENCE OF 417-1356 FROM N.A.
RC TISSUE=Fibrinolytic;
RA Saito M., Kunisaki N., Hirono I., Aoki T., Ishida M., Urano N.,
RA Kimura S.;
RT "Partial characterization of cDNA clones encoding the three distinct
RT pro alpha chains of type I collagen from rainbow trout."
RL Fisheries Sci. 64:780-786(1998).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB052837; BAB55663.1; -
DR EMBL; AB008372; BAA33379.1; -
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01410; COLF1.1.
DR Pfam; PF01391; COLLAGEN; 17.
DR Prodom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLF1.1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
DR Glycoprotein; Collagen; Signal.
DR SIGNAL; 1 24
DR PROPEP 25 ? AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
DR CHAIN ? 1096 COLLAGEN ALPHA 2(I) CHAIN.
DR PROPEP 1097 1356 CARBOXYL-TERMINAL PROPEPTIDE
DR (BY SIMILARITY).
DR CARBOHYD 1257 1257 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR SEQUENCE 1356 AA; 126985 MW; 7BB2F1F80DB10C93 CRC64;
SQ
Query Match 62.0%; Score 57; DB 1; Length 1356;
Best Local Similarity 71.4%; Pred. No. 0.71;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
OY 1 GVGKDGKGNPGMPGA 14
DB 937 GFKGDRGPGSPGA 950

```

```

RA Liakka A., Thesleff I., Kraal G., Tryggvason K.;
RT "Cloning of a novel bacteria-binding receptor structurally related to
RT scavenger receptors and expressed in a subset of macrophages."
RL Cell 80:603-609(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9265975; PubMed=10331948;
RA Kangas M., Brannstrom A., Elomaa O., Matsuda Y., Eddy R., Shows T.B.,
RA Tryggvason K.;
RT "Structure and chromosomal localization of the human and murine genes
RT for the macrophage MARCO receptor."
RL Genomics 58:82-89(1999).
CC -1- FUNCTION: Bind Gram-positive and Gram-negative bacteria.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
CC -1- TISSUE SPECIFICITY: Expressed in subpopulations of macrophages in
CC the spleen and the mediastinal lymph nodes.
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U18424; AAA68638.1; -
DR EMBL; AF128423; AAD51136.1; -
DR EMBL; AF127927; AAD51136.1; JOINED.
DR EMBL; AF127928; AAD51136.1; JOINED.
DR EMBL; AF128169; AAD51136.1; JOINED.
DR EMBL; AF128170; AAD51136.1; JOINED.
DR EMBL; AF128171; AAD51136.1; JOINED.
DR EMBL; AF127601; AAD51136.1; JOINED.
DR EMBL; AF127602; AAD51136.1; JOINED.
DR EMBL; AF128419; AAD51136.1; JOINED.
DR EMBL; AF128420; AAD51136.1; JOINED.
DR EMBL; AF128421; AAD51136.1; JOINED.
DR EMBL; AF128422; AAD51136.1; JOINED.
DR MGI; MGI:1309998; Marco.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001190; SRCR.
DR Pfam; PF01391; COLLAGEN; 4.
DR Pfam; PF00530; SRCR; 1.
DR PRINTS; PR00258; SPERACTRCPT.
DR SMART; SM00202; SR; 1.
DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
DR PROSITE; PS0287; SRCR_2; 1.
DR Collagen; Transmembrane; Receptor; Glycoprotein; Signal-anchor.
DR KW Collagen; Transmembrane; Receptor; Glycoprotein; Signal-anchor.
DR TRANSMEM 49 69
DR DOMAIN 1 48
DR DOMAIN 149 518
DR DOMAIN 423 518
DR DISULFID 446 507 BY SIMILARITY.
DR DISULFID 459 517 BY SIMILARITY.
DR DISULFID 487 497 BY SIMILARITY.
DR CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR SEQUENCE 518 AA; 52730 MW; B09E7601ECA23637 CRC64;
SQ
Query Match 60.9%; Score 56; DB 1; Length 518;
Best Local Similarity 66.7%; Pred. No. 0.39;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 1 GVGKDGKGNPGMPGAP 15
DB 294 GVGKDGKPGVGVP 308

```

RESULT 28

CALL\_HUMAN STANDARD: PRT: 1143 AA.  
ID CALL\_HUMAN 014993: 013676; 012885; 09H572; 09NP2; 09NP22;  
AC 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 1(XIX) chain precursor (Collagen alpha 1(Y) chain).  
GN COL19A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
ON NCBI\_TaxID=9606;  
RX MEDLINE=95293914; PubMed=775380;  
RA Inoguchi K., Yoshiohka H., Khaleduzzaman M., Ninomiya Y.;  
RT "The mRNA for alpha 1(XIX) collagen chain, a new member of FACITs,  
RT contains a long unusual 3' untranslated region and displays many  
RT unique splicing variants.";  
RL J. Biochem. 117:137-146(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=98008918; PubMed=9344653;  
RA Khaleduzzaman M., Sumiyoshi H., Ueki Y., Inoguchi K., Ninomiya Y.,  
RA Yoshiohka H.;  
RT "Structure of the human type XIX collagen (COL19A1) gene, which  
RT suggests it has arisen from an ancestor gene of the FACIT family.";  
RL Genomics 45:304-312(1997).  
RN [3]  
RP SEQUENCE OF 132-952 FROM N.A.  
RA Blakey S., Parker A., Skuce C.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 739-1143 FROM N.A.  
RC TISSUE=Skin;  
RA MEDLINE=94308092; PubMed=8034603;  
RA Myers J.C., Yang H., D'Ippolito J.A., Presente A., Miller M.K.,  
RA Dion A.S.;  
RT "The triple-helical region of human type XIX collagen consists of  
RT multiple collagenous subdomains and exhibits limited sequence homology  
RT to alpha 1(XVI).";  
RL J. Biol. Chem. 269:18549-18557(1994).  
CC -1- FUNCTION: May act as a cross-bridge between fibrils and other  
CC extracellular matrix molecules.  
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (By  
CC similarity).  
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
CC INTERRUPTED HELICES (FACIT) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: D38163; BAA07368.1; -  
DR EMBL: AB004629; BAA23309.1; -  
DR EMBL: AB004634; BAA23309.1; JOINED.  
DR EMBL: AB004604; BAA23309.1; JOINED.  
DR EMBL: AB004615; BAA23309.1; JOINED.  
DR EMBL: AB004626; BAA23309.1; JOINED.  
DR EMBL: AB004630; BAA23309.1; JOINED.  
DR EMBL: AB004631; BAA23309.1; JOINED.  
DR EMBL: AB004632; BAA23309.1; JOINED.  
DR EMBL: AB004633; BAA23309.1; JOINED.  
DR EMBL: AB004634; BAA23309.1; JOINED.  
DR EMBL: AB004584; BAA23309.1; JOINED.  
DR EMBL: AB004585; BAA23309.1; JOINED.  
DR EMBL: AB004586; BAA23309.1; JOINED.  
DR EMBL: AB004587; BAA23309.1; JOINED.  
DR EMBL: AB004588; BAA23309.1; JOINED.

DR EMBL: AB004589; BAA23309.1; JOINED.  
DR EMBL: AB004590; BAA23309.1; JOINED.  
DR EMBL: AB004591; BAA23309.1; JOINED.  
DR EMBL: AB004592; BAA23309.1; JOINED.  
DR EMBL: AB004593; BAA23309.1; JOINED.  
DR EMBL: AB004594; BAA23309.1; JOINED.  
DR EMBL: AB004595; BAA23309.1; JOINED.  
DR EMBL: AB004596; BAA23309.1; JOINED.  
DR EMBL: AB004597; BAA23309.1; JOINED.  
DR EMBL: AB004598; BAA23309.1; JOINED.  
DR EMBL: AB004599; BAA23309.1; JOINED.  
DR EMBL: AB004600; BAA23309.1; JOINED.  
DR EMBL: AB004601; BAA23309.1; JOINED.  
DR EMBL: AB004602; BAA23309.1; JOINED.  
DR EMBL: AB004603; BAA23309.1; JOINED.  
DR EMBL: AB004605; BAA23309.1; JOINED.  
DR EMBL: AB004606; BAA23309.1; JOINED.  
DR EMBL: AB004607; BAA23309.1; JOINED.  
DR EMBL: AB004608; BAA23309.1; JOINED.  
DR EMBL: AB004609; BAA23309.1; JOINED.  
DR EMBL: AB004610; BAA23309.1; JOINED.  
DR EMBL: AB004611; BAA23309.1; JOINED.  
DR EMBL: AB004612; BAA23309.1; JOINED.  
DR EMBL: AB004613; BAA23309.1; JOINED.  
DR EMBL: AB004614; BAA23309.1; JOINED.  
DR EMBL: AB004616; BAA23309.1; JOINED.  
DR EMBL: AB004617; BAA23309.1; JOINED.  
DR EMBL: AB004618; BAA23309.1; JOINED.  
DR EMBL: AB004619; BAA23309.1; JOINED.  
DR EMBL: AB004620; BAA23309.1; JOINED.  
DR EMBL: AB004621; BAA23309.1; JOINED.  
DR EMBL: AB004622; BAA23309.1; JOINED.  
DR EMBL: AB004623; BAA23309.1; JOINED.  
DR EMBL: AB004624; BAA23309.1; JOINED.  
DR EMBL: AB004625; BAA23309.1; JOINED.  
DR EMBL: AB004627; BAA23309.1; JOINED.  
DR EMBL: AL118519; CAB99331.1; -  
DR EMBL: AL133388; CAC01297.1; -  
DR EMBL: AL136445; CAC12699.1; -  
DR EMBL: U09279; AAA21146.1; -  
DR EMBL: U09280; AAA21147.1; -  
DR EMBL: U09281; AAA21147.1; JOINED.  
DR MIM: 120165; -  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR003129; TSPN.  
DR Pfam: PF01391; Collagen; 12.  
DR Pfam: PF02210; TSPN; 1.  
DR SMART: SM00210; TSPN; 1.  
DR EMBL: U09281; AAA21147.1; JOINED.  
DR MIM: 120165; -  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR003129; TSPN.  
DR Pfam: PF01391; Collagen; 12.  
DR Pfam: PF02210; TSPN; 1.  
DR SMART: SM00210; TSPN; 1.  
KW Cell adhesion; Collagen; Signal.  
FT SIGNAL 1 23  
FT CHAIN 24 1143  
FT DOMAIN 50 235  
FT DOMAIN 293 352  
FT DOMAIN 371 430  
FT DOMAIN 449 689  
FT DOMAIN 701 819  
FT DOMAIN 834 1013  
FT DOMAIN 1055 1112  
FT SITE 953 955  
FT SITE 953 955  
FT CONFLICT 89 90  
FT CONFLICT 280 280  
FT CONFLICT 366 366  
FT CONFLICT 442 443  
FT CONFLICT 938 938  
FT CONFLICT 1141 1141  
SQ SEQUENCE 1143 AA; 115448 MW; 0F1A0D825589B378 CRC64;  
Query Match 60.9%; Score 56; DB 1; Length 1143;  
Best Local Similarity 66.7%; Pred. No. 0.85;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

Db      457 GIKGDKGEGTGLPGFP 471
      ||||| |||
RESULT 29
CAL3_MOUSE STANDARD: PRT; 1464 AA.
AC P08121; Q61429; Q9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RL complete DNA sequence."
RN [2]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogel G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RL alpha-1 type-III collagen chain."
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Lian G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RL (III) collagen gene."
RN [4]
RP SEQUENCE OF 810-1464 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic head;
RL MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aikawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado T., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Futuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RN [5]
RP SEQUENCE OF 1442-1464 FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RL collagen mRNAs."
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.

```

```

CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLINES.
CC -1- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCANS CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 WVEC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52046; CAA36279.1; -;
DR EMBL; M18933; AAA37338.1; -;
DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR EMBL; AK019448; BAB31724.1; -;
DR EMBL; X57983; CAA41048.1; -;
DR PIR; A22287; A22287.
DR PIR; A27353; A27353.
DR PIR; S16373; S16373.
DR MGD; MGI:88453; Col3a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR01007; WVEC.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF002078; Fib_collagen_C; 1.
DR PRODOM; PD002078; Fib_collagen_C; 1.
DR SMART; SM0038; COLFI; 1.
DR SMART; SM00214; WVEC; 1.
DR PROSITE; PS01208; WVEC; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 154
FT CHAIN 155 1203
FT PROPEP 1204 1464
FT DOMAIN 31 90
FT DOMAIN 155 169
FT DOMAIN 170 1195
FT DOMAIN 1196 1464
FT CARBOHYD 262 262
FT MOD_RES 262 262
FT MOD_RES 283 283
FT MOD_RES 859 859
FT MOD_RES 976 976
FT MOD_RES 1093 1093
FT MOD_RES 1105 1105
FT DISULFD 1195 1195
FT DISULFD 1196 1196
SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;
Query Match 60.9%; Score 56; DB 1; Length 1464;
Best Local Similarity 69.2%; Pred. No. 1.1;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Ox 1 GYKGDGKGNPGMPG 13
Db 857 GYKGDGKGNPGMPG 869

```

```

RESULT 30
MSRE_MOUSE STANDARD: PRT; 458 AA.
ID MSRE_MOUSE
AC P30204; Q9QZ56;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Macrophage scavenger receptor types I and II (Macrophage acetylated
DE LDL receptor I and II) (Scavenger receptor type A) (SR-A).

```

GN MSRI OR SCVR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBITaxID=10080;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS I AND II).  
 RA MEDLINE=93359822; PubMed=8394868;  
 RA Ashkenas J., Penman M., Vassile E., Accon S., Freeman M.W.,  
 RT Krueger M.;  
 RT Structures and high and low affinity ligand binding properties of  
 RT murine type I and type II macrophage scavenger receptors.";  
 RL J. Lipid Res. 34:983-1000(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM II).  
 RA MEDLINE=93313972; PubMed=8380589;  
 RA Doi T., Wada Y., Kodama T., Higashi K.I., Kurihara Y.,  
 RA Miyazaki T., Nakamura H., Uesugi S., Imanishi T., Kawabe Y.,  
 RA Itakura H., Yazaki Y., Matsumoto A.;  
 RT "Charged collagen structure mediates the recognition of negatively  
 RT charged macromolecules by macrophage scavenger receptors.";  
 RL J. Biol. Chem. 268:2126-2133(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM I).  
 RC STRAIN=C57BL/6;  
 RA Rateri D.L., Whitman S.C., Block A.E., Daugherty A.;  
 RT Identification of a functional domain in class A scavenger receptors  
 RT that mediates metabolism of AcLDL.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 349-458 FROM N.A.  
 RX MEDLINE=91062370; PubMed=1978939;  
 RA Freeman M., Ashkenas J., Rees D.J., Kingsley D.M., Copeland N.G.,  
 RA Jenkins N.A., Krueger M.;  
 RT "An ancient, highly conserved family of cysteine-rich protein domains  
 RT revealed by cloning type I and type II murine macrophage scavenger  
 RT receptors.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).  
 RN [5]  
 RP SEQUENCE OF 1-4 FROM N.A.  
 RX MEDLINE=95395388; PubMed=7666008;  
 RA Altring R.P., Freeman M.W.;  
 RT Structure of the murine macrophage scavenger receptor gene and  
 RT evaluation of sequences that regulate expression in the macrophage  
 RT cell line, P388D.";  
 RL J. Lipid Res. 36:1305-1314(1995).  
 CC -1- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC  
 CC DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.  
 CC TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDATE THE  
 CC ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING  
 CC MODIFIED LOW DENSITY LIPOPROTEINS (LDL).  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L04274; AAA39747.1; ALT\_INIT.  
 CC EMBL: L04275; AAA39748.1; ALT\_INIT.  
 CC EMBL: D13382; BAA02650.1; -.  
 CC EMBL: AF203781; AAF14001.1; -.  
 CC EMBL: M59445; AAA37464.1; -.  
 CC EMBL: M59446; AAA37465.1; -.  
 CC EMBL: U13873; AAC13774.1; -.  
 CC PIR: A38260; A38260.

DR MGI:98257; Msrl.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001190; SRCR.  
 DR Pfam: PF01391; Collagen; 1.  
 DR Pfam: PF00330; SRCR; 1.  
 DR PRINTS: PR00258; SPERACTRCPTR.  
 DR SMART: SM00202; SR; 1.  
 DR PROSITE: PS00420; SRCR\_1; 1.  
 DR PROSITE: PS0287; SRCR\_2; 1.  
 KW Transmembrane; Glycoprotein; Endocytosis; Colled coll; LDL;  
 KW Receptor; Signal-anchor; Alternative splicing.  
 FT DOMAIN 1 55  
 FT TRANSMEM 56 78  
 FT FT  
 FT DOMAIN 79 458  
 FT DOMAIN 79 114  
 FT DOMAIN 209 259  
 FT DOMAIN 277 350  
 FT DOMAIN 357 457  
 FT CARBOHYD 94 94  
 FT CARBOHYD 107 107  
 FT CARBOHYD 147 147  
 FT CARBOHYD 188 188  
 FT CARBOHYD 253 253  
 FT CARBOHYD 271 271  
 FT CARBSPLIC 352 354  
 FT VARSPLIC 355 458  
 FT CONFLICT 47 47  
 FT CONFLICT 110 110  
 FT CONFLICT 120 120  
 FT CONFLICT 130 130  
 FT CONFLICT 168 168  
 FT CONFLICT 202 202  
 SQ SEQUENCE 458 AA; 50130 MW; 6D4C041C27EE50B CRC64;  
 Query Match 60.3%; Score 55.5; DB 1; Length 458;  
 Best Local Similarity 50.0%; Pred. No. 0.41;  
 Matches 12; Conservative 1; Mismatches 2; Indels 9; Gaps 1;  
 QY 1 GYKGD-----KGNPGNPGAP 15  
 DB 307 GYKGDRCQIGFPGCRGNPGAPKRP 330

Search completed: November 1, 2002, 12:52:53  
 Job time : 9 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:19:21 ; Search time 21.5 seconds  
(without alignments)  
120.694 Million cell updates/sec

Title: US-09-529-691A-1  
Perfect score: 92  
Sequence: 1 GVGKDGKNGMPCAP 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_19: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	1621	4	Q9H4R9
2	69	73.0	886	4	Q9NUB7
3	66	71.7	404	2	Q9F687
4	64	69.6	673	4	Q14052
5	64	69.6	1691	11	Q9ESQ2
6	63	68.5	174	13	Q90585
7	63	68.5	979	13	Q919K3
8	62	67.4	290	5	Q26054
9	62	67.4	412	9	Q9F2Y3
10	62	67.4	1414	5	Q26634
11	62	67.4	1670	4	Q9B0T2
12	62	67.4	1802	5	Q17163
13	61	66.3	224	5	Q9VYK2
14	61	66.3	447	4	Q16593
15	61	66.3	957	4	Q9H0V3
16	61	66.3	957	4	Q96P44

17	60	65.2	310	13	Q90612	Q90612 gallus gall
18	60	65.2	1497	4	Q9NQK9	Q9NQK9 homo sapien
19	60	65.2	1497	4	Q9UWD9	Q9UWD9 homo sapien
20	60	65.2	1532	4	Q02802	Q02802 homo sapien
21	60	65.2	1761	5	Q18407	Q18407 drosophila
22	60	65.2	1940	5	Q9VNV5	Q9VNV5 drosophila
23	60	65.2	2944	11	Q63870	Q63870 mus musculu
24	59.5	64.7	907	5	Q26312	Q26312 strongyloce
25	59.5	64.7	1752	5	Q07265	Q07265 strongyloce
26	59	64.7	771	4	Q9UJC7	Q9UJC7 homo sapien
27	59	64.1	1431	11	Q9JMH4	Q9JMH4 mesocricetu
28	59	64.1	1433	11	Q07563	Q07563 mus musculu
29	59	64.1	1682	11	Q9QZK9	Q9QZK9 mus musculu
30	59	64.1	1737	11	Q9J104	Q9J104 rattus norv
31	58	63.0	410	2	Q9F691	Q9F691 streptococc
32	58	63.0	464	13	Q90412	Q90412 brachydantlo
33	58	63.0	654	4	Q9BXS0	Q9BXS0 homo sapien
34	58	63.0	666	11	Q99MQ5	Q99MQ5 mus musculu
35	58	63.0	713	5	Q9GV24	Q9GV24 sarcophaga
36	58	63.0	1691	11	Q9ESQ1	Q9ESQ1 mus musculu
37	58	63.0	1758	5	Q19098	Q19098 caenorhabd1
38	58	63.0	1759	5	Q19099	Q19099 caenorhabd1
39	57	62.0	132	4	P78429	P78429 homo sapien
40	57	62.0	278	4	Q9BX19	Q9BX19 homo sapien
41	57	62.0	650	5	Q17866	Q17866 caenorhabd1
42	57	62.0	684	5	P90679	P90679 arenicola m
43	57	62.0	778	5	Q9U9K6	Q9U9K6 caenorhabd1
44	57	62.0	864	5	Q93336	Q93336 caenorhabd1
45	57	62.0	890	5	Q77087	Q77087 alvinella p

## ALIGNMENTS

RESULT 1  
ID Q9H4R9 PRELIMINARY: PRT: 1621 AA.  
AC Q9H4R9;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE BA472K17.2 (COLLAGEN TYPE IV ALPHA 1) (FRAGMENT).  
GN COL4A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bates K.;  
RL Submitted (OCCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL390755; CAC13153.1; -.  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER  
SQ SEQUENCE 1621 AA: 155705 MW: 73F6F901CD0EDBA2 CRC64;  
FT NON\_TER  
SQ SEQUENCE 1621 AA: 155705 MW: 73F6F901CD0EDBA2 CRC64;  
Query Match 100.0%; Score 92; DB 4; Length 1621;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GVGKDGKNGMPCAP 15  
DB 1215 GVGKDGKNGMPCAP 1229  
RESULT 2  
ID Q9NUB7 PRELIMINARY: PRT: 886 AA.  
AC Q9NUB7;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE BA472K17.2 (COLLAGEN TYPE IV ALPHA 1) (FRAGMENT).  
GN COL4A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bates K.;  
RL Submitted (OCCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL390755; CAC13153.1; -.  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER  
SQ SEQUENCE 1621 AA: 155705 MW: 73F6F901CD0EDBA2 CRC64;  
FT NON\_TER  
SQ SEQUENCE 1621 AA: 155705 MW: 73F6F901CD0EDBA2 CRC64;  
Query Match 100.0%; Score 92; DB 4; Length 1621;  
Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC 09NDB7;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE DA4AA3.1 (COLLAGEN, TYPE IV, ALPHA 5 (ALPORT SYNDROME))  
DE (FRAGMENT).  
GN COL4A5.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Cobley V.;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AL035425; CAB90289.1; -  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR Hum. GeneL. 77:318-324(1987).  
DR EMBL: M24766; AAA52043.1; -  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR ProdDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER 1 1  
SQ SEQUENCE 886 AA; 85479 MW; 8C06B9FCA9AA6569 CRC64;  
Query Match 75.0%; Score 69; DB 4; Length 886;  
Best Local Similarity 73.3%; Pred. No. 0.027;  
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKDGNPMPGAP 15  
DB 480 GIKGKGNPMPGAP 494  
ID 09F687 PRELIMINARY; PRT; 404 AA.  
AC 09F687;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-AP34;  
RC MEDLINE=20490571; PubMed=11035747;  
RA Rasmussen M., Eden A., Bjorck L.;  
RT "Sc1a, a novel collagen-like surface protein of streptococcus pyogenes.";  
RL Infect. Immun. 68:6370-6377(2000).  
DR EMBL: AF296336; AAG30216.1; -  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 3.  
FT NON\_TER 1 1  
FT NON\_TER 404 404  
SQ SEQUENCE 404 AA; 41612 MW; 45304C89E3B41A78 CRC64;  
Query Match 71.7%; Score 66; DB 2; Length 404;  
Best Local Similarity 73.3%; Pred. No. 0.034;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GVKDGNPMPGAP 15  
DB 113 GIKGKGNPMPGAP 127  
RESULT 4  
014052

ID 014052 PRELIMINARY; PRT; 673 AA.  
AC 014052;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE ALPHA-2 TYPE IV COLLAGEN (FRAGMENT).  
GN COL4A2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RX MEDLINE=88085168; PubMed=3692475;  
RA Killen P.D., Francmano C.A., Yamada Y., Modi W.S., O'Brien S.J.;  
RT "Partial structure of the human alpha 2(IV) collagen chain and chromosomal localization of the gene (COL4A2).";  
RL Hum. GeneL. 77:318-324(1987).  
DR EMBL: M24766; AAA52043.1; -  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR ProdDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Collagen.  
FT NON\_TER 1 1  
SQ SEQUENCE 673 AA; 67174 MW; D2F3C9B3111A3105 CRC64;  
Query Match 69.6%; Score 64; DB 4; Length 673;  
Best Local Similarity 73.3%; Pred. No. 0.12;  
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKDGNPMPGAP 15  
DB 329 GPKGKGNPMPGAP 343  
ID 09ESQ2 PRELIMINARY; PRT; 1691 AA.  
AC 09ESQ2;  
DT 01-MAR-2001 (TREMblrel. 16, Created)  
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
DE TYPE IV COLLAGEN ALPHA 5 CHAIN.  
GN COL4A5.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20536494; PubMed=10965041;  
RA Saito K., Naito I., Seki T., Ohashi T., Kimura E., Momota R., Kishiro Y., Sado Y., Yoshioke H., Ninomiya Y.;  
RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in Epithelial Basement Membranes.";  
RL J. Biochem. 128:427-434(2000).  
DR EMBL: AB041350; BAB13673.1; -  
DR InterPro: IPR001442; C4.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01413; C4; 2.  
DR ProdDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;  
Query Match 69.6%; Score 64; DB 11; Length 1691;  
Best Local Similarity 66.7%; Pred. No. 0.3;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;



Oy 1 GVGKDGKNGPCGP 15  
 Db 1285 GVGKDGKNGPCGP 1299

## RESULT 6

ID 090585 PRELIMINARY; PRT: 174 AA.  
 AC 090585;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE CARTRIDGE ALPHA-1(IX) COLLAGEN-PROTEOGLYCAN (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90062114; PubMed=2584206;  
 RA Nishimura I., Muragaki Y., Olsen B.R.;  
 RT "Tissue-specific forms of type IX collagen-proteoglycan arise from the  
 use of two widely separated promoters.";  
 RL J. Biol. Chem. 264:20033-20041(1989).  
 DR EMBL: M28660; AAA48709.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 KW COLLAGEN.  
 FT NON\_TER  
 SQ SEQUENCE 174 AA; 16240 MW; C0E64A0748A224F0 CRC64;

Query Match 68.5%; Score 63; DB 13; Length 174;  
 Best Local Similarity 66.7%; Pred. No. 0.041;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GVGKDGKNGPCGP 15  
 Db 46 GVGKDGKNGPCGP 60

## RESULT 7

ID 0919K3 PRELIMINARY; PRT: 979 AA.  
 AC 0919K3;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE COLLAGEN IV A1 CHAIN (FRAGMENT).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Halford W.M., Dong S.;  
 RT "Composition, synthesis and assembly of the embryonic chick retinal  
 basal lamina.";  
 RL Dev. Biol. 0:0-0(2000).  
 DR EMBL: AF239838; AAF44681.1; -;  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 12.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 FT NON\_TER  
 SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 68.5%; Score 63; DB 13; Length 979;  
 Best Local Similarity 60.0%; Pred. No. 0.24;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GVGKDGKNGPCGP 15  
 Db 573 GVGKDGKNGPCGP 587

## RESULT 8

ID 026054 PRELIMINARY; PRT: 290 AA.  
 AC 026054;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE COLLAGEN-LIKE PROTEIN (FRAGMENT).  
 OS Paracentrotus lividus (Common sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinidae;  
 OC Paracentrotus.  
 OX NCBI\_TaxID=7656;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89149773; PubMed=2537631;  
 RA Salta B., Butlice G., Gambino R.;  
 RT "Isolation of a putative collagen-like gene from the sea urchin  
 Paracentrotus lividus.";  
 RL Biochem. Biophys. Res. Commun. 158:633-639(1989).  
 DR EMBL: M24558; AAA29439.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 4.  
 FT NON\_TER  
 SQ SEQUENCE 290 AA; 26174 MW; FA221A1BF1E67D2F CRC64;

Query Match 67.4%; Score 62; DB 5; Length 290;  
 Best Local Similarity 78.6%; Pred. No. 0.098;  
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GVGKDGKNGPCGP 14  
 Db 76 GVGKDGKNGPCGP 89

## RESULT 9

ID 09FZY3 PRELIMINARY; PRT: 412 AA.  
 AC 09FZY3;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE ORF62.  
 OS bacteriophage phi ETA.  
 OC Viruses.  
 OX NCBI\_TaxID=106284;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20566787; PubMed=1115106;  
 RA Nakayama K., Yamada S., Komatsuzawa H., Sugai M.;  
 RT "Phage conversion of exfoliative toxin A production in *Staphylococcus aureus*.";  
 RL Mol. Microbiol. 38:694-705(2000).  
 DR EMBL: AP001553; BAA97648.1; -;  
 DR InterPro: IPR000087; Collagen.  
 SQ SEQUENCE 412 AA; 46013 MW; 3E87F7CBAE223088 CRC64;

Query Match 67.4%; Score 62; DB 9; Length 412;  
 Best Local Similarity 78.6%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GVGKDGKNGPCGP 14

```

Db      178 GAKGDKGPGQPGA 191
      ||||| ||||
RESULT 10
026634      PRELIMINARY;      PRT; 1414 AA.
AC      026634;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE      ALPHA-1 COLLAGEN.
GN      COL1A1PRA.
OS      Strongylocentrotus purpuratus (Purple sea urchin).
OC      Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC      Echinoidae; Euechinoidea; Echinacea; Echinoidae; Strongylocentrotidae;
OC      Strongylocentrotus.
OX      NCBI_TaxID=7668;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92348411; Pubmed=1639795;
RA      Exposito J.-Y., D'Alessio M., Solursh M., Ramirez F.;
RT      "Sea urchin collagen evolutionarily homologous to vertebrate pro-
      alpha-2(I) collagen".
RL      J. Biol. Chem. 267:15559-15562(1992).
DR      EMBL; M92040; AAA30035.1; -.
DR      InterPro: IPR000087; Collagen.
DR      InterPro: IPR000885; Fib-collagen_C.
DR      Pfam: PF01410; COLFI; 1.
DR      Pfam: PF01391; Collagen; 18.
DR      ProDom: PD002078; Fib-collagen_C; 1.
DR      SMART; SM00038; COLFI; 1.
KW      Collagen.
SQ      SEQUENCE 1414 AA; 133025 MW; 94D9CA71A9FD73D CRC64;

Query Match      67.4%; Score 62; DB 5; Length 1414;
Best Local Similarity 78.6%; Pred. No. 0.51;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 GVKGDKNPMPGA 14
      | ||| |||| |
Db      393 GSKGDGNGPQPGA 406

RESULT 11
09BOT2      PRELIMINARY;      PRT; 1670 AA.
AC      09BOT2;
DT      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT      01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE      ALPHA3 TYPE IV COLLAGEN.
GN      COL4A3.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=21064696; Pubmed=11134255;
RA      Heidet L., Attondel C., Forestier L., Cohen-Solal L., Mollet G.,
      Ramirez B., Stavrou C., Gubler M.C., Antignac C.;
RT      "Structure of the human type IV collagen gene COL4A3 and mutations in
      the autosomal Alport syndrome.".
RL      J. Am. Soc. Nephrol. 12:97-106(2001).
DR      EMBL; AJ288487; CAC36101.1; -.
DR      EMBL; AJ288488; CAC36101.1; JOINED.
DR      EMBL; AJ288489; CAC36101.1; JOINED.
DR      EMBL; AJ288490; CAC36101.1; JOINED.
DR      EMBL; AJ288491; CAC36101.1; JOINED.
DR      EMBL; AJ288492; CAC36101.1; JOINED.
DR      EMBL; AJ288493; CAC36101.1; JOINED.
DR      EMBL; AJ288494; CAC36101.1; JOINED.
DR      EMBL; AJ288495; CAC36101.1; JOINED.

```

```

DR      EMBL; AJ288496; CAC36101.1; JOINED.
DR      EMBL; AJ288497; CAC36101.1; JOINED.
DR      EMBL; AJ288498; CAC36101.1; JOINED.
DR      EMBL; AJ288499; CAC36101.1; JOINED.
DR      EMBL; AJ288500; CAC36101.1; JOINED.
DR      EMBL; AJ288501; CAC36101.1; JOINED.
DR      EMBL; AJ288502; CAC36101.1; JOINED.
DR      EMBL; AJ288503; CAC36101.1; JOINED.
DR      EMBL; AJ288504; CAC36101.1; JOINED.
DR      EMBL; AJ288505; CAC36101.1; JOINED.
DR      EMBL; AJ288506; CAC36101.1; JOINED.
DR      EMBL; AJ288507; CAC36101.1; JOINED.
DR      EMBL; AJ288508; CAC36101.1; JOINED.
DR      EMBL; AJ288509; CAC36101.1; JOINED.
DR      EMBL; AJ288510; CAC36101.1; JOINED.
DR      EMBL; AJ288511; CAC36101.1; JOINED.
DR      EMBL; AJ288512; CAC36101.1; JOINED.
DR      EMBL; AJ288513; CAC36101.1; JOINED.
DR      EMBL; AJ288514; CAC36101.1; JOINED.
DR      EMBL; AJ288515; CAC36101.1; JOINED.
DR      EMBL; AJ288516; CAC36101.1; JOINED.
DR      EMBL; AJ288517; CAC36101.1; JOINED.
DR      EMBL; AJ288518; CAC36101.1; JOINED.
DR      EMBL; AJ288519; CAC36101.1; JOINED.
DR      EMBL; AJ288520; CAC36101.1; JOINED.
DR      EMBL; AJ288521; CAC36101.1; JOINED.
DR      EMBL; AJ288522; CAC36101.1; JOINED.
DR      EMBL; AJ288523; CAC36101.1; JOINED.
DR      EMBL; AJ288524; CAC36101.1; JOINED.
DR      EMBL; AJ288525; CAC36101.1; JOINED.
DR      EMBL; AJ288526; CAC36101.1; JOINED.
DR      EMBL; AJ288527; CAC36101.1; JOINED.
DR      EMBL; AJ288528; CAC36101.1; JOINED.
DR      EMBL; AJ288529; CAC36101.1; JOINED.
DR      EMBL; AJ288530; CAC36101.1; JOINED.
DR      EMBL; AJ288531; CAC36101.1; JOINED.
DR      EMBL; AJ288532; CAC36101.1; JOINED.
DR      EMBL; AJ288533; CAC36101.1; JOINED.
DR      EMBL; AJ288534; CAC36101.1; JOINED.
DR      EMBL; AJ288535; CAC36101.1; JOINED.
DR      EMBL; AJ288536; CAC36101.1; JOINED.
DR      EMBL; AJ288537; CAC36101.1; JOINED.
DR      EMBL; AJ288538; CAC36101.1; JOINED.
DR      InterPro: IPR001442; C4.
DR      InterPro: IPR000087; Collagen.
DR      InterPro: IPR000504; RRM.
DR      InterPro: IPR001563; Serine_carppept.
DR      Pfam: PF01413; C4; 2.
DR      Pfam: PF01391; Collagen; 21.
DR      ProDom: PD003923; C4; 2.
DR      SMART; SM00111; C4; 2.
DR      PROSITE; PS00131; CARBOXYPEPT_SER_SER; UNKNOWN_1.
DR      PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW      Collagen.
SQ      SEQUENCE 1670 AA; 161899 MW; FA7BE4914CA0A6F6 CRC64;

Query Match      67.4%; Score 62; DB 4; Length 1670;
Best Local Similarity 66.7%; Pred. No. 0.6;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY      1 GVKGDKNPMPGAP 15
      | ||| | |||| |
Db      1189 GAKGDRGAPGFGP 1203

RESULT 12
017163      PRELIMINARY;      PRT; 1802 AA.
AC      017163;
DT      01-NOV-1996 (TREMBlrel. 01, Created)
DT      01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBlrel. 19, Last annotation update)

```

```

DE A2 (IV) BASEMENT MEMBRANE COLLAGEN.
OS Brugia malayi.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioides;
OC Onchocercidae; Brugia.
OX NCBI_TaxID=6279;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=APERIODIC;
RX MEDLINE=95364846; PubMed=7637709;
RA Cautel V.R., Rajan T.V.
RT "The structural organization of an alpha 2 (type IV) basement membrane
RT collagen gene from the filarial nematode Brugia malayi."
RL Mol. Biochem. Parasitol. 70:227-229(1995).
DR EMBL: U07224; AAC46611.1; -.
DR InterPro: IPR001442; C4.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 23.
DR ProDom: PD003923; C4; 2.
DR SMART: SM00111; C4; 2.
KM Collagen.
SQ SEQUENCE 1802 AA; 172401 MW; 595F16554CBF2D24 CRC64;

Query Match 67.4%; Score 62; DB 5; Length 1802;
Best Local Similarity 71.4%; Pred. No. 0.65;
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPCGPGCA 14
DB 1196 GIRGDKGSPGIPGA 1209
I:::||||:|||||

RESULT 13
O9VVY2 PRELIMINARY; PRT; 224 AA.
ID O9VVY2
AC O9VVY2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG14089 PROTEIN.
GN CG14089.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceolnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Burtis K.C., Burtis D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merkulov G., Milshina N.V., Modary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003517; AAF49173.1; -.
DR FlyBase: Fgn0036861; CG14089.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR002965; P-rich_extensn.
DR Pfam: PF01391; Collagen; 1.
DR PRINTS: PR01217; PRICHEXTENS.
SQ SEQUENCE 224 AA; 24308 MW; F84B9C912D8EC1CD CRC64;

Query Match 66.3%; Score 61; DB 5; Length 224;
Best Local Similarity 66.7%; Pred. No. 0.11;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPCGPGAP 15
DB 118 GPRGDKGHPGPGIP 132
I:::||||:|||||

RESULT 14
O16593 PRELIMINARY; PRT; 447 AA.
ID O16593
AC O16593;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COLLAGEN-LIKE PROTEIN (447 AA) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Kimura S.;
RL Submitted (FEB-1990) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Kimura S.;
RL Submitted (MAY-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: X14963; CAA33085.1; -.
DR EMBL: X15038; CAA33142.1; -.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF01391; Collagen; 5.
FT NON_TER 1
FT NON_TER 447
SQ SEQUENCE 447 AA; 41829 MW; FDB207023D87CC94 CRC64;

Query Match 66.3%; Score 61; DB 4; Length 447;
Best Local Similarity 73.3%; Pred. No. 0.22;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 GVGKDGKNGPCGPGAP 15
DB 280 GOKDGKNGPCDPTP 294
I:::||||:|||||

RESULT 15
O9H0V3

```

ID Q9H0V3 PRELIMINARY; PRT; 957 AA.  
AC Q9H0V3;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 99.4 KDA PROTEIN.  
GN DKF2P564B052.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wieman S., Weil B., Wellenreuther R., Gassenhuber J., Gissel S.,  
Ansoorge W., Boecker M., Bloecker H., Bauersachs S., Blum H.,  
Lauder J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
Mewes H.W., Ottenwälder B., Obermayer B., Tampe J., Heubner D.,  
Wambutt R., Korn B., Klein M., Poustka A.;  
RA "Towards a Catalog of Human Genes and Proteins: Sequencing and  
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.";  
RL Genome Res. 11:422-435(2001).  
DR EMBL; AL136624; CAB6559.1; -.  
DR HSSP; P17301; 1A0X.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003129; TSPN.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01391; Collagen; 6.  
DR Pfam; PF02210; TSPN; 1.  
DR Pfam; PF00092; vwa; 1.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00327; VWFA; 1.  
DR PROSITE; PS50234; VWFA; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 957 AA; 99413 MW; 391022EF715024571 CRC64;

Query Match 66.3%; Score 61; DB 4; Length 957;  
Best Local Similarity 66.7%; Pred. No. 0.48;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKDGKGNPWPAP 15  
ID 1 GVKDGKGNPWPAP 15  
AC Q96P44; PRELIMINARY; PRT; 957 AA.  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE COLLAGEN XXI.  
GN COL21A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Fitzgerald J., Bateman J.F.;  
RT "A new FACIT of the collagen family: COL21A1.";  
RL FEBS Lett. 0:0-0(2001).  
DR EMBL; AF414088; AAL02227.1; -.  
SQ SEQUENCE 957 AA; 99369 MW; 4C5CDF5E6656A675 CRC64;

Query Match 66.3%; Score 61; DB 4; Length 957;  
Best Local Similarity 66.7%; Pred. No. 0.48;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKDGKGNPWPAP 15

Db 551 GAKGKGNAGPGLP 565  
ID 090612 PRELIMINARY; PRT; 310 AA.  
AC Q90612;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 27.6 KDA PROTEIN (FRAGMENT).  
GN COL3A1.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94266842; PubMed=8206952;  
RA Nah H.D., Niu Z., Adams S.L.;  
RT "An alternative transcript of the chick type III collagen gene that  
RT does not encode type III collagen.";  
RL J. Biol. Chem. 269:16443-16448(1994).  
DR EMBL; U07974; AAA83409.1; -.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 5.  
KW Hypothetical protein.  
FT NON\_TER 310 310  
SQ SEQUENCE 310 AA; 27601 MW; 5C60B4360832814C CRC64;

Query Match 65.2%; Score 60; DB 13; Length 310;  
Best Local Similarity 71.4%; Pred. No. 0.21;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDGKGNPWPAP 14  
ID 1 GVKDGKGNPWPAP 14  
AC Q9NOK9; PRELIMINARY; PRT; 1497 AA.  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE BA16H23.2 (COLLAGEN, TYPE XVII, ALPHA 1 (BP180)).  
GN COL17A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Brown A.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL138761; CAC00589.1; -.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 7.  
KW Collagen.  
SQ SEQUENCE 1497 AA; 150419 MW; E01027005F3AB843 CRC64;

Query Match 65.2%; Score 60; DB 4; Length 1497;  
Best Local Similarity 78.6%; Pred. No. 1.1;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKDGKGNPWPAP 14  
ID 1 GVKDGKGNPWPAP 14  
Db 960 GPKDGKGNPWPAP 973

RESULT	19
09UMD9	PRELIMINARY; PRT: 1497 AA.
ID	09UMD9
AC	09UMD9
DT	01-MAY-2000 (TREMBLER, 13, Created)
DT	01-MAY-2000 (TREMBLER, 13, Last sequence update)
DT	01-JUN-2001 (TREMBLER, 17, last annotation update)
DE	180 KDA BULLOUS PEMPHIGOID ANTIGEN 2/TYPE XVII COLLAGEN.
GN	BPA62/COL17A1.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_Textid:9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE-9716601; PubMed-9012408;
RA	Gatalica B., Pulkkinen L., Li K., Kuokkanen K., Ryyanen M.,
RA	McGrath J.A., Ultio J.;
RT	"Cloning of the human type XVII collagen gene (COL17A1), and detection
RT	of novel mutations in generalized atrophic benign epidermolysis
RT	bullosa.";
RL	Am. J. Hum. Genet. 60:352-365(1997).
DR	EMBL; U76604; AAB51499.1; JOINED.
DR	EMBL; U76565; AAB51499.1; JOINED.
DR	EMBL; U76566; AAB51499.1; JOINED.
DR	EMBL; U76567; AAB51499.1; JOINED.
DR	EMBL; U76568; AAB51499.1; JOINED.
DR	EMBL; U76569; AAB51499.1; JOINED.
DR	EMBL; U76570; AAB51499.1; JOINED.
DR	EMBL; U76571; AAB51499.1; JOINED.
DR	EMBL; U76572; AAB51499.1; JOINED.
DR	EMBL; U76573; AAB51499.1; JOINED.
DR	EMBL; U76574; AAB51499.1; JOINED.
DR	EMBL; U76575; AAB51499.1; JOINED.
DR	EMBL; U76576; AAB51499.1; JOINED.
DR	EMBL; U76577; AAB51499.1; JOINED.
DR	EMBL; U76578; AAB51499.1; JOINED.
DR	EMBL; U76579; AAB51499.1; JOINED.
DR	EMBL; U76580; AAB51499.1; JOINED.
DR	EMBL; U76581; AAB51499.1; JOINED.
DR	EMBL; U76582; AAB51499.1; JOINED.
DR	EMBL; U76583; AAB51499.1; JOINED.
DR	EMBL; U76584; AAB51499.1; JOINED.
DR	EMBL; U76585; AAB51499.1; JOINED.
DR	EMBL; U76586; AAB51499.1; JOINED.
DR	EMBL; U76587; AAB51499.1; JOINED.
DR	EMBL; U76588; AAB51499.1; JOINED.
DR	EMBL; U76589; AAB51499.1; JOINED.
DR	EMBL; U76590; AAB51499.1; JOINED.
DR	EMBL; U76591; AAB51499.1; JOINED.
DR	EMBL; U76592; AAB51499.1; JOINED.
DR	EMBL; U76593; AAB51499.1; JOINED.
DR	EMBL; U76594; AAB51499.1; JOINED.
DR	EMBL; U76595; AAB51499.1; JOINED.
DR	EMBL; U76596; AAB51499.1; JOINED.
DR	EMBL; U76597; AAB51499.1; JOINED.
DR	EMBL; U76598; AAB51499.1; JOINED.
DR	EMBL; U76599; AAB51499.1; JOINED.
DR	EMBL; U76600; AAB51499.1; JOINED.
DR	EMBL; U76601; AAB51499.1; JOINED.
DR	EMBL; U76602; AAB51499.1; JOINED.
DR	EMBL; U76603; AAB51499.1; JOINED.
DR	InterPro: IPR000087; Collagen.
DR	Pfam: PF01391; Collagen; 6.
SV	SEQUENCE
SV	1497 AA; 150459 MW; 952E107C6DA6B5C CRC64;
Query Match	65.2%; Score 60; DB 4; Length 1497;
Best Local Similarity	78.6%; Pred. No. 1.1;
Matches	11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
1	GVKGDGNPMPGA 14
1	:

```

Db      960  GPKGDKGDPGVPGA  973

RESULT 20
ID      002802      PRELIMINARY;      PRT: 1532 AA.
AC      002802;
DT      01-NOV-1996 (TREMBLrel. 01, Created)
DT      01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      AUTOANTIGEN (FRAGMENT).
GN      Bp180.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=FORESKIN;
RX      MEDLINE=92381323; PubMed=1324962;
GX      Giudice G.J., Emery D.J., Diaz L.A.;
RT      "Cloning and primary structural analysis of the Bullous pemphigoid
RT      autoantigen, Bp180."
RL      J. Invest. Dermatol. 99:243-250(1992).
DR      EMBL; M9169; AAA35605.1; -.
DR      InterPro; IPR000087; Collagen.
DR      Pfam; PF01391; Collagen; 6.
FT      NON-TER
FT      SEQUENCE 1532 AA; 154568 MW; BF35054CF93B8EE3 CRC64;

Oy      1  GVKGDKGNPGWPGAP  14
      1  |||||||:||||
Db      995  GPKGDKGDPGVPGA  1008

RESULT 21
ID      018407      PRELIMINARY;      PRT: 1761 AA.
AC      018407;
DT      01-JAN-1998 (TREMBLrel. 05, Created)
DT      01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      COLLAGEN TYPE IV ALPHA 2.
VX      VKG OR DMCOLN2 OR CG16858.
GN      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
CX      NCBI_TaxID=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Yasothornrikul S., Davis W.J., Cramer G., Kimbrell D.A.,
RA      Dearolf C.R.;
RL      Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR      EMBL; U65431; AAB64082.1; -.
DR      FlyBase; FBgn0016075; VKG.
DR      InterPro; IPR001442; C4.
DR      Pfam; PF01413; C4; 2.
DR      Pfam; PF01391; Collagen; 20.
DR      ProDom; PD003923; C4; 2.
DR      SMART; SM00111; C4; 2.
SO      SEQUENCE 1761 AA; 175955 MW; FC823AEFF19121DC6 CRC64;

Query Match      65.2%; Score 60; DB 4; Length 1532;
Best Local Similarity 78.6%; Pred. No. 1.1;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy      1  GVKGDKGNPGWPGAP  15

```

Db 994 GRKGTGNPFGPGRP 1008

## RESULT 22

Q9VWV5 PRELIMINARY; PRT; 1940 AA.  
ID Q9VWV5;  
AC Q9VWV5;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE VKG PROTEIN.  
GN VKG OR CG16858.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Colniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.T., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Butcher K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Chertys J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwa C.,  
RA Kimmel B.E., Knelson F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mezuklov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskaes R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weissenhach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003608; AAF52203.1; -;  
DR FlyBase; FBgn0016075; VKG.  
DR InterPro; IPR001442; C4.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR PRODOM; PD003923; C4; 2.  
DR SMART; SM00111; C4; 2.  
SQ SEQUENCE 1940 AA; 193777 MW; 9B507382EF9C17B5 CRC64;

Query Match 65.2%; Score 60; DB 5; Length 1940;  
Best Local Similarity 66.7%; Pred. NO. 1.4;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GVKDGKNGPFGAP 15  
Db 996 GRKGTGNPFGPGRP 1010

## RESULT 23

063870 PRELIMINARY; PRT; 2944 AA.  
ID 063870;  
AC 063870;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TYPE VII COLLAGEN.  
GN COL7A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93315168; PubMed=8325648;  
RA Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L.,  
RA Jenkins N.A., Uitto J.;  
RT "CDNA cloning and chromosomal mapping of the mouse type VII collagen  
RT gene (Col7a1): evidence for rapid evolutionary divergence of the  
RL gene.";  
RL Genomics 16:733-739(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96275720; PubMed=8752674;  
RA Kivirikko S., Li K., Christiano A.M., Uitto J.;  
RT "Cloning of mouse type VII collagen reveals evolutionary conservation  
RT of functional protein domains and genomic organization.";  
RL J. Invest. Dermatol. 106:1300-1306(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Kivirikko S., Li K., Christiano A.M., Uitto J.;  
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.  
DR EMBL; U32107; AAB6593.1; -;  
DR EMBL; S63654; AAB27492.1; -;  
DR HSSP; P12111; 2KNT.  
DR MGI; 88462; Col7a1.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR002223; Kunitz\_BPTI.  
DR InterPro; IPR002035; VWFA.  
DR Pfam; PF01391; Collagen; 22.  
DR Pfam; PF00041; Fn3; 9.  
DR Pfam; PF00014; Kunitz\_BPTI; 1.  
DR Pfam; PF00092; vwa; 2.  
DR PRINTS; PR00759; BASICPASE.  
DR PRINTS; PR00453; VWFADOMAIN.  
DR SMART; SM00060; FN3; 9.  
DR SMART; SM00327; VWA; 2.  
DR PROSITE; PS00280; BPTI\_KUNITZ\_1; 1.  
DR PROSITE; PS0279; BPTI\_KUNITZ\_2; 1.  
DR PROSITE; PS50234; VWFA; 2.  
KV Serine protease inhibitor.  
SQ SEQUENCE 2944 AA; 295116 MW; 596FA507BC6C02C2 CRC64;

Query Match 65.2%; Score 60; DB 11; Length 2944;  
Best Local Similarity 73.3%; Pred. NO. 2.2;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GVKDGKNGPFGAP 15  
Db 2377 GVKGDLGPFGAP 2391

RESULT 24  
Q26312 PRELIMINARY; PRT; 907 AA.  
ID Q26312

AC 026312;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE NONFIBRILLAR COLLAGEN PROTEIN (FRAGMENT).  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92038439; PubMed=1936564;  
 RA Messel G.M., Elkin M., Benson S.;  
 RT "Primary mesenchyme cells of the sea urchin embryo require an  
 RT autonomously produced, nonfibrillar collagen for spiculogenesis.";  
 RL Dev. Biol. 148:261-272(1991).  
 DR EMBL: S64572; AAB20270.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 13.  
 KW Collagen.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 907 AA; 85400 MW; B8ECBF6FDA59D88 CRC64;  
 Query Match 64.7%; Score 59.5; DB 5; Length 907;  
 Best Local Similarity 54.2%; Pred. No. 0.77;  
 Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

OY 1 GVKGD-----KGNGWPGAP 15  
 Db 546 GVKGDGPGRTGPEGAKGNPGLPGIP 569

RESULT 25  
 007265 PRELIMINARY; PRT; 1752 AA.  
 ID 007265;  
 AC 007265;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE 3 ALPHA PROCOLLAGEN.  
 GN COL3ALPHA.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93186842; PubMed=8444899;  
 RA Exposito J.-Y., D'Alessio M., Di Liberto M., Ramirez F.;  
 RT "Complete primary structure of a sea-urchin type IV collagen and  
 RT analysis of the 5' end of its gene.";  
 RL J. Biol. Chem. 268:5249-5254(1993).  
 DR EMBL: L02917; AAA30039.1; -;  
 DR InterPro: IPR001442; C4.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01413; C4; 2.  
 DR Pfam: PF01391; Collagen; 22.  
 DR ProDom: PD003923; C4; 2.  
 DR SMART: SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1752 AA; 170210 MW; 1AE5AAA21569346D CRC64;  
 Query Match 64.7%; Score 59.5; DB 5; Length 1752;  
 Best Local Similarity 54.2%; Pred. No. 1.5;  
 Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

OY 1 GVKGD-----KGNGWPGAP 15  
 Db 1177 GVKGDGPGRTGPEGAKGNPGLPGIP 1200

RESULT 26  
 09JUC7 PRELIMINARY; PRT; 771 AA.  
 ID 09JUC7;  
 AC 09JUC7;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE COL4A5 (COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR) (FRAGMENT).  
 GN COL4A5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Bird C.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AL031622; CAA20937.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 10.  
 KW Collagen.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 771 AA; 72760 MW; 9CFCBD373E545A61 CRC64;  
 Query Match 64.1%; Score 59; DB 4; Length 771;  
 Best Local Similarity 73.3%; Pred. No. 0.78;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKNPGWPGAP 15  
 Db 528 GVKGDKNPGWPGAP 542

RESULT 27  
 09JMH4 PRELIMINARY; PRT; 1431 AA.  
 ID 09JMH4;  
 AC 09JMH4;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE COLLAGEN TYPE XVII.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;  
 RT "Mesocricetus auratus mRNA for type XVII collagen.";  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB027759; BAA94381.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 5.  
 KW Collagen.  
 SQ SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;  
 Query Match 64.1%; Score 59; DB 11; Length 1431;  
 Best Local Similarity 66.7%; Pred. No. 1.5;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKNPGWPGAP 15  
 Db 945 GVKGDGDPGVPGTP 959

RESULT 28  
 007563 PRELIMINARY; PRT; 1433 AA.  
 ID 007563;  
 AC 007563;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:14:41 : Search time 26.5 Seconds  
(without alignments)  
62,872 Million cell updates/sec

Title: US-09-529-691A-3  
Perfect score: 92  
Sequence: 1 PAGPMGPNKGDKVG 15

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	68.5	534	18	AAW12844
2	63	68.5	535	18	AAW12841
3	63	68.5	537	18	AAW12840
4	63	68.5	674	21	AAW53439
5	63	68.5	1040	21	AAW84547
6	63	68.5	1040	21	AAW84548
7	63	68.5	1366	16	AAW71702
8	63	68.5	1366	21	AAW96123
9	63	68.5	1366	21	AAW56801
10	63	68.5	1366	22	AAW50293
11	63	68.5	1366	22	AAW02536

12	59	64.1	1418	15	AAW59751
13	59	64.1	1418	22	AAW35624
14	57	62.0	1442	16	AAW79480
15	56	60.9	97	21	AAW44187
16	55	59.8	54	16	AAW79481
17	55	59.8	1418	16	AAW71703
18	55	59.8	1418	21	AAW96124
19	55	59.8	1487	19	AAW61562
20	54	58.7	1078	16	AAW71704
21	54	58.7	1078	21	AAW96125
22	54	58.7	1196	13	AAW28916
23	54	58.7	1288	18	AAW26328
24	54	58.7	1288	20	AAW92297
25	54	58.7	1466	22	AAW50291
26	54	58.7	1466	22	AAW02533
27	54	58.7	1466	22	AAW02534
28	54	58.7	1466	22	AAW02537
29	54	58.7	1469	22	AAW02537
30	54	58.7	1838	15	AAW53257
31	53	57.6	337	22	AAW75419
32	53	57.6	470	22	AAW84421
33	53	57.6	471	22	AAW84419
34	53	57.6	501	22	AAW84420
35	53	57.6	936	22	AAW70107
36	53	57.6	1745	22	AAW77793
37	52.5	57.1	532	19	AAW50033
38	52	56.5	36	22	AAW39385
39	52	56.5	36	22	AAW24171
40	52	56.5	36	22	AAW60063
41	52	56.5	36	22	AAW72668
42	52	56.5	36	22	AAW32900
43	52	56.5	200	22	AAW02714
44	52	56.5	200	22	AAW68068
45	52	56.5	219	21	AAW84553

#### ALIGNMENTS

RESULT 1  
ID AAW12844 standard; peptide: 534 AA.  
XX AAW12844:  
AC  
XX  
DT 15-DEC-1997 (first entry)  
XX  
DE Pro-alpha2(I):((III)BGR chimeric protein.  
XX  
KW C-propeptide: recognition sequence; procollagen: monomer chain; therapy;  
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
KW fibrotic disease; human; chimeric protein.  
XX  
OS Homo sapiens.  
XX  
PN W09708311-A1.  
XX  
PD 06-MAR-1997.  
XX  
PF 30-AUG-1996; 96MO-GB02122.  
XX  
PR 14-JUN-1996; 96GB-0012476.  
PR 31-AUG-1995; 95GB-0017773.  
PR 23-MAR-1996; 96GB-0006152.  
PA (UYMA-) UNIV VICTORIA MANCHESTER.  
XX  
PI Bullfield N, Kadler K;  
XX  
DR WPI, 1997-179268/16.  
XX  
PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached to an allen collagen alpha-chain or non-collagen material,

PT useful e.g. for wound healing  
XX  
PS Example 1; Page 38-41; 69pp; English.  
XX  
CC This sequence represents a chimeric procollagen molecule of the  
CC invention. This sequence has the procollagen C-propeptide from the  
CC pro-alpha1(III) chain inserted into the pro-alpha2(I) chain sequence. The  
CC C-propeptide is implicated in the assembly of the monomer chains into  
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
CC and formation of collagen in fibril-forming pro-alpha chains. The  
CC C-propeptides determine the type-specific assembly of the moieties to  
CC which they are attached. The molecule of the invention comprises a first  
CC moiety having procollagen C-propeptide activity attached to a second  
CC moiety, which is an alien collagen molecule can be used for treatment of  
CC material. The novel collagen molecule can be used for treatment of  
CC diagnosis in humans or animals, especially for the treatment of  
CC wound healing or fibrotic diseases with reduced scarring or for use in  
CC photocoagulation, brewing, foodstuffs or textiles. The novel collagen  
CC molecules, especially when containing substitutions in the recognition  
CC site, may have significantly altered properties and characteristics, such  
CC as different binding kinetics or alpha-chain selection properties.  
XX  
SQ Sequence 534 AA:  
Query Match 68.5%; Score 63; DB 18; Length 534;  
Best Local Similarity 73.3%; Pred. No. 0.65;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 1 PAGPMGPNKDKRGV 15  
||||| 11 ||||| 1  
DB 223 PAGSGPAGKDKRGTG 237  
RESULT 2  
AAW12841  
ID AAW12841 standard; Protein; 535 AA.  
XX  
AC AAW12841;  
XX  
DT 15-DEC-1997 (first entry)  
XX  
DE Truncated pro-alpha2(I) chain.  
XX  
KM C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
KW fibrotic disease; human.  
XX  
OS Homo sapiens.  
XX  
PN WO9708311-A1.  
XX  
PD 06-MAR-1997.  
XX  
PF 30-AUG-1996; 96WO-GB02122.  
XX  
PR 14-JUN-1996; 96GB-0012476.  
PR 31-AUG-1995; 95GB-0017773.  
PR 23-MAR-1996; 96GB-0006152.  
XX  
PA (UYMA-) UNIV VICTORIA MANCHESTER.  
XX  
PI Bulleid N, Kadler K;  
XX  
DR WPI: 1997-179268/16.  
DR N-PSDB; AAT59891.  
XX  
PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide  
PT attached to an alien collagen alpha-chain or non-collagen material,  
XX useful e.g. for wound healing  
XX  
PS Example 1; Page 25-28; 69pp; English.

CC This sequence represents a truncated procollagen pro-alpha2(I) chain that  
CC can be used in the procollagen molecules of the invention. The  
CC C-propeptide is implicated in the assembly of the monomer chains into  
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
CC and formation of collagen in fibril-forming pro-alpha chains. The  
CC C-propeptides determine the type-specific assembly of the moieties to  
CC which they are attached. The molecule of the invention comprises a first  
CC moiety having procollagen C-propeptide activity attached to a second  
CC moiety, which is an alien collagen molecule can be used for treatment of  
CC material. The novel collagen molecule can be used for treatment of  
CC diagnosis in humans or animals, especially for the treatment of  
CC wound healing or fibrotic diseases with reduced scarring or for use in  
CC photocoagulation, brewing, foodstuffs or textiles. The novel collagen  
CC molecules, especially when containing substitutions in the recognition  
CC site, may have significantly altered properties and characteristics, such  
CC as different binding kinetics or alpha-chain selection properties.  
XX  
SQ Sequence 535 AA:  
Query Match 68.5%; Score 63; DB 18; Length 535;  
Best Local Similarity 73.3%; Pred. No. 0.65;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 1 PAGPMGPNKDKRGV 15  
||||| 11 ||||| 1  
DB 224 PAGSGPAGKDKRGTG 238  
RESULT 3  
AAW12840  
ID AAW12840 standard; peptide; 537 AA.  
XX  
AC AAW12840;  
XX  
DT 15-DEC-1997 (first entry)  
XX  
DE Pro-alpha2(I):(III) CP chimeric protein.  
XX  
KM C-propeptide; recognition sequence; procollagen; monomer chain; therapy;  
KW trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;  
KW fibrotic disease; human; chimeric protein.  
XX  
OS Homo sapiens.  
XX  
PN WO9708311-A1.  
XX  
PD 06-MAR-1997.  
XX  
PF 30-AUG-1996; 96WO-GB02122.  
XX  
PR 14-JUN-1996; 96GB-0012476.  
PR 31-AUG-1995; 95GB-0017773.  
PR 23-MAR-1996; 96GB-0006152.  
XX  
PA (UYMA-) UNIV VICTORIA MANCHESTER.  
XX  
PI Bulleid N, Kadler K;  
XX  
DR WPI: 1997-179268/16.  
XX  
PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide  
PT attached to an alien collagen alpha-chain or non-collagen material,  
XX useful e.g. for wound healing  
XX  
PS Claim 16; Page 35-38; 69pp; English.  
XX  
CC This sequence represents a chimeric procollagen molecule of the  
CC invention. This sequence has the procollagen C-propeptide from the  
CC pro-alpha1(III) chain attached to the pro-alpha2(I) chain sequence. The  
CC C-propeptide is implicated in the assembly of the monomer chains into  
CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
CC and formation of collagen in fibril-forming pro-alpha chains. The



CC proline. The method is also useful in studying the structure and function  
CC of polypeptides which do not normally contain trans-4-hydroxyproline.  
CC The present sequence represents human collagen 1 (alpha2) helical region,  
CC which may be produced using the method of the invention.

XX Sequence 1040 AA;

Query Match 68.5%; Score 63; DB 21; Length 1040;

Best Local Similarity 73.3%; Pred. No. 1.3;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPWGPNKGDKVG 15

Db 976 PAGPSGPAKGDKRTG 990

RESULT 6  
AAY84548  
ID AAY84548 standard; Protein; 1040 AA.

XX AAY84548;

XX 25-JUL-2000 (first entry)

XX A human collagen 1 (alpha2) protein helical region.

XX Extracellular matrix protein; self aggregation; hydroxylated proline;

XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;

XX collagen; fibrinogen; fibronectin; post translational hydroxylation.

XX Synthetic.

XX Homo sapiens.

XX EP992586-A2.

XX 12-APR-2000.

XX 07-OCT-1999; 99EP-0119184.

XX 09-OCT-1998; 98US-0169768.

XX (USSU ) US SURGICAL CORP.

XX Gruskin EA, Buechter DD, Zhang G, Connolly K;

XX WPI: 2000-259138/23.

XX N-PSDB: AAA12511.

XX Production of extracellular matrix proteins containing

XX 4-trans-hydroxyproline results in native self aggregating proteins,

XX useful on medical implants -

XX Example 14; Fig 50A-E; 260pp; English.

XX The specification describes a method for producing an extracellular

XX matrix protein or its fragment. The extracellular matrix protein is

XX capable of self aggregating in a cell which does not ordinarily

XX hydroxylated prolines. The method comprises optimising a nucleic acid

XX sequence for expression in the cell by substitution of codons preferred

XX by that cell for naturally occurring codons not preferred by the cell;

XX incorporating the nucleic acid sequence into the cell; and contacting

XX the cell with a hypertonic growth medium containing at least one amino

XX acid, selected from the group consisting of trans-4-hydroxyproline and

XX 3-hydroxyproline to allow at least one of the amino acids to be

XX assimilated into the cell and incorporated into the extracellular matrix

XX protein. The method may be used to make host cells assimilate and

XX incorporate trans-4-hydroxyproline into proteins. This is especially

XX useful in the recombinant production of proteins such as collagen,

XX fibrinogen and fibronectin whose ability to self aggregate and produce

XX functional proteins depends on the post translational hydroxylation of

XX proline. The method is also useful in studying the structure and function

XX of polypeptides which do not normally contain trans-4-hydroxyproline.

XX The present sequence represents a human collagen 1 (alpha2) helical

CC region, encoded by a sequence which has optimised codon usage. The

CC protein may be produced using the method of the invention.

XX Sequence 1040 AA;

Query Match 68.5%; Score 63; DB 21; Length 1040;

Best Local Similarity 73.3%; Pred. No. 1.3;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPWGPNKGDKVG 15

Db 976 PAGPSGPAKGDKRTG 990

RESULT 7  
AAR71702  
ID AAR71702 standard; protein; 1366 AA.

XX AAR71702;

XX 17-OCT-1995 (first entry)

XX Collagen alpha 2 (I) chain precursor.

XX Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;

XX disorder; osteoporosis; metastatic progression; Paget's disease;

XX hyperthyroidism; bone; resorption; rheumatoid arthritis;

XX osteoarthritis; vasculitis syndrome.

XX Homo sapiens.

XX WO9508115-A.

XX 23-MAR-1995.

XX 19-SEP-1994; 94WO-DK00348.

XX 17-SEP-1993; 93DK-0001040.

XX (OSTE-) OSTEOMETER AS.

XX Bonde M, Qvist P;

XX WPI: 1995-131456/17.

XX Assaying collagen fragments in body fluid by immunoassay - using

XX antibodies raised against synthetic peptide(s) contg. potential

XX crosslinking sites, to diagnose and monitor disorders of collagen

XX metabolism, e.g. osteoporosis.

XX Disclosure (Appendix A); Page 51; 87pp; English.

XX Determination of collagen fragments in body fluids can be achieved

XX by immunoassay using antibodies directed against synthetic peptides

XX derived from collagen which contain sites of potential crosslinking.

XX The method is used to diagnose and monitor treatment of disorders of

XX collagen metabolism (degradation of type I collagen may indicate

XX osteoporosis, metastatic progression, Paget's disease,

XX hyperthyroidism or other conditions involving excessive bone

XX resorption; degradation of type II collagen may indicate rheumatoid

XX arthritis or osteoarthritis; and of type III collagen, vacuolitis

XX syndrome). The method can also be used to assess the toxicity of a

XX compound and to test drugs for their effect on collagen metabolism.

XX Sequence 1366 AA;

Query Match 68.5%; Score 63; DB 16; Length 1366;

Best Local Similarity 73.3%; Pred. No. 1.7;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPWGPNKGDKVG 15

Db 1055 PAGPSGPAKGDKRTG 1069

```

RESULT 8
AAV96123
ID AAV96123 standard; Peptide; 1366 AA.
XX
XX AAV96123;
AC
XX
XX 19-DEC-2000 (first entry)
DT
XX
XX Collagen type I alpha-2.
DE
XX
XX Collagen type I; osteoporosis; bone resorption; Paget's disease;
KM hyperparathyroidism; metastasis; assay; diagnosis.
XX
OS Homo sapiens.
XX
XX US6110689-A.
XX
XX 29-AUG-2000.
PD
XX
XX 04-NOV-1997; 97US-0963825.
PF
XX
XX 21-JAN-1994; 94US-0187319.
PR
XX
XX (OSTE-) OSTEOMETER AS.
PA
XX
XX Bonde M, Qvist P;
PI
XX
XX WPI; 2000-586349/55.
DR
XX
XX Assaying type I collagen fragments for diagnosing osteoporosis in
PT postmenopausal women, involves contacting body fluid with synthetic
PT collagen peptide and antibody and quantifying by competitive binding
PT assay -
XX
XX Disclosure; Column 31-37; 41pp; English.
XX
XX The present sequence is that of human type I collagen alpha-2.
CC The invention is based on the discovery of the presence of
CC particular collagen fragments in body fluids of patients compared
CC with those of healthy subjects. These fragments are generated
CC upon collagen degradation and are partly characterized by the
CC presence of potential sites for crosslinking. A method for
CC assaying collagen fragments in a body fluid sample is based on the
CC competitive binding to immunological binding partners of collagen
CC fragments in the sample and of synthetic peptides derived from
CC collagen and containing crosslinkable sites (see AAV96105-11). When
CC considering the degradation of type I collagen, the assay can be
CC used as a means of identifying excessive bone resorption, indicating
CC the presence of osteoporosis or the metastatic progress of a
CC malignancy. Other conditions characterized by excessive bone
CC resorption include Paget's disease and hyperparathyroidism.
XX
XX
SQ Sequence 1366 AA:

```

```

Query Match 68.5%; Score 63; DB 21; Length 1366;
Best Local Similarity 73.3%; Pred. No. 1.7;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 PAGPWGPNKGKGVG 15
DB 1055 PAGPSPGAKGKGRG 1069

```

```

RESULT 9
AAV56801
ID AAV56801 standard; protein; 1366 AA.
XX
XX AAV56801;
AC
XX
XX 27-MAR-2000 (first entry)
DT
XX

```

```

DE Human preproalpha 2 (I) collagen.
XX
XX Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
KM sequence selection and alignment domain; prosthetic implant;
KM foodstuff; medicine; type I collagen; human.
XX
XX Homo sapiens.
OS
XX
XX EP967226-A2.
XX
XX 29-DEC-1999.
PD
XX
XX 04-MAY-1999; 99EP-0303470.
PF
XX
XX 08-MAY-1998; 98US-0084828.
PR
XX
XX 10-APR-1999; 99US-0289578.
XX
XX (COHE-) COHESION TECHNOLOGIES INC.
XX
XX Olsen DR, Hitzeman RA, Chisholm GE;
XX
XX WPI; 2000-074666/07.
DR
XX
XX New method for production of fibrillar collagen, useful for preparing
PT telopeptide collagen fibrils and gelatin -
PT
XX
XX Example 1; Fig 4A-B; 30pp; English.
PS
XX
XX The invention provides a method for the production of fibrillar collagen.
CC The method comprises: (a) culturing a recombinant host cell comprising a
CC DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
CC (sequence selection and alignment domain); and (b) producing the
CC fibrillar collagen. The methods are used to produce fibrillar collagen,
CC from which telopeptide collagen fibrils can be derived. Host cells,
CC comprising DNA encoding a collagen monomer lacking SSAD or N propeptide
CC is used to produce gelatin. Collagen is used in biological research
CC as a substrate for in vitro cell culture and as a component of
CC biocompatible materials for use in prosthetic implants, sustained drug
CC release matrices, artificial skin and wound dressing and healing devices.
CC Gelatin is particularly useful for foodstuffs and medicine, for coating
CC tablets and making capsules. The methods, comprising the use of collagen
CC monomers lacking the N and/or C propeptides, result in a large increase
CC in the production of type I collagen. The present sequence represents
CC the human preproalpha 2 (I) collagen (Genbank Accn no: Z74616).
XX
XX
SQ Sequence 1366 AA:

```

```

Query Match 68.5%; Score 63; DB 21; Length 1366;
Best Local Similarity 73.3%; Pred. No. 1.7;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 PAGPWGPNKGKGVG 15
DB 1055 PAGPSPGAKGKGRG 1069

```

```

RESULT 10
ABB50293
ID ABB50293 standard; Protein; 1366 AA.
XX
XX ABB50293;
AC
XX
XX 08-FEB-2002 (first entry)
DT
XX
XX Collagen type I alpha-2 ovarian tumour marker protein, SEQ ID NO:76.
DE
XX
XX Ovarian tumour marker gene; human; overexpression; upregulation;
KM epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
KM identification; serous cystadenoma; borderline serous tumour;
KM serous cystadenocarcinoma; mucinous cystadenocarcinoma;
KM mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
KM undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
KM adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;

```



```

XX  Collagen: triple helix; articular cartilage; collagenase;
KM  degradation; monoclonal antibody; epitope; matrix;
XX  metalloproteinase.
OS  Homo sapiens.
XX  Key
FH  Location/Qualifiers
FT  Peptide 1..24
FT  /label= Signal peptide.
XX  WO9414070-A.
XX  23-JUN-1994.
XX  06-DEC-1993: 93WO-CA00522.
XX  04-DEC-1992: 92US-0984123.
XX  (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.
XX  Hollander AP, Poole AR;
XX  WPI: 1994-234222/28.
XX  Determ. of cartilage degradation - using a monoclonal antibody
PT  to measure the amt. of unwound collagen or fragments in samples
XX  Disclosure: Figure 1: 119pp; English.
XX  Type II collagen constitutes the bulk of the fibrillar backbone of
CC  cartilage matrix. It is composed of a tightly wound triple helix
CC  which can only be cleaved by the metalloproteinase collagenase to
CC  produce 3/4 and 1/4 length alpha chain fragments. The destruction of
CC  articular cartilage is due, in part, to the degradation of collagen.
CC  Inceperable of maintaining its helical structure at physiological
CC  temperatures, collagenase-cleaved collagens unwind and become
CC  susceptible to further degradation by other proteinases. By
CC  producing monoclonal antibodies directed against epitopes which are
CC  only revealed when collagen is unwound, the antibodies provide a
CC  means of determining the degradation of cartilage in a biological
CC  sample. The antibodies do not bind to native helical collagen.
CC  Epitopes used in the production of such antibodies are described in
CC  AAR59749, AAR59750 and AAR57742.
XX  SQ  Sequence 1418 AA:
XX  Query Match 64.1%; Score 59; DB 15; Length 1418;
XX  Best Local Similarity 66.7%; Pred. No. 6.3;
XX  Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
XX  QY 1 PAGPMPGNGKRGKVG 15
XX  Db 1096 PGPVPGSPGKDGAMG 1110
XX  RESULT 13
XX  AAB35624
XX  ID AAB35624 standard; Protein; 1418 AA.
XX  AC AAB35624;
XX  DT 14-FEB-2001 (first entry)
XX  DE Human type II collagen.
XX  KW Type II collagen; arthritis; joint; ds.
XX  OS Homo sapiens.
XX  PN US6132976-A.
XX  PD 17-OCT-2000.

```

```

XX  22-JAN-1998: 98US-0010999.
XX  04-DEC-1992: 92US-0984123.
XX  17-JUL-1993: 95US-0448501.
XX  (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.
XX  Billinghamst RC, Poole AR, Hollander AP;
XX  WPI: 2001-006136/01.
XX  Detecting cartilage degradation useful for early detection of arthritis
PT  or joint damage by contacting the biological sample with an antibody
PT  that binds to an epitope of unwound type II collagen chains but not to
PT  a native helical collagen.
XX  Example 1: Fig 1: 58pp; English.
XX  The present invention relates to detecting cartilage degradation in a
CC  biological sample by identifying the presence of unwound type II
CC  collagen in the sample. The method involves contacting the sample with
CC  a monoclonal antibody which only binds an epitope on unwound type II
CC  collagen chains. The invention is useful for the early detection of
CC  arthritis and joint damage and for monitoring disease related to
CC  collagen.
XX  SQ  Sequence 1418 AA:
XX  Query Match 64.1%; Score 59; DB 22; Length 1418;
XX  Best Local Similarity 66.7%; Pred. No. 6.3;
XX  Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
XX  QY 1 PAGPMPGNGKRGKVG 15
XX  Db 1096 PGPVPGSPGKDGAMG 1110
XX  RESULT 14
XX  AAR79480
XX  ID AAR79480 standard; Protein; 1442 AA.
XX  AC AAR79480;
XX  DT 19-JAN-1996 (first entry)
XX  DE Rat type II collagen.
XX  KW Collagen: bone progenitor; gene transfer; gene therapy; osteoporosis;
XX  osteotomy; bone repair; osteotropic; Pichia pastoris.
XX  OS Rattus sp.
XX  PN WO9522611-A2.
XX  PD 24-AUG-1995.
XX  21-FEB-1995: 95WO-US02251.
XX  30-SEP-1994: 94US-0316650.
XX  18-FEB-1994: 94US-0199780.
XX  (UNMT ) UNIV MICHIGAN.
XX  DR WPI: 1995-302717/39.
XX  Transferring nucleic acid into bone progenitor cell(s) - using a
PT  bone compatible matrix, for treatment of fracture(s) and
PT  osteoporosis.
XX  Disclosure: Page 197-208; 317pp; English.
XX  Human, rat and mouse collagen type II (given in AAR79479-81,
XX  CC

```

CC respectively) can be used to stimulate bone progenitor cells as a  
CC means of treating bone-related diseases in association with an  
CC osteotropic gene.  
XX  
SQ Sequence 1442 AA;

Query Match 62.0%; Score 57; DB 16; Length 1442;  
Best Local Similarity 73.3%; Pred. No. 12;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 1 PAGPWGPNKDGKVG 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 871 PGPVGSKDGDPKG 885

RESULT 15  
AAB44187  
ID AAB44187 standard; Protein: 97 AA.

AC AAB44187;  
DT 08-FEB-2001 (first entry)  
XX

DE Human cancer associated protein sequence SEQ ID NO:1632.

XX Human; cancer associated gene; cancer antigen; detection; cancer;  
KW diagnosis; cytostatic; proliferative; vulnary; immunomodulator;  
KW antidiabetic; antistatic; antirheumatic; antiallergic; antiviral;  
KW antiinflammatory; antihypertensive; antibacterial; cardiac;  
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
KW vasotropic; antipruritic; antidiabetic; gene therapy; inflammation;  
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KW allergic reaction; graft versus host disease; organ rejection;  
KW haemostatic; thrombolytic; cardiovascular disorder; infection;  
KW neurological disease; drug screening.

XX Homo sapiens.

OS WO200055350-A1.

PN 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI: 2000-587533/55.

DR N-PSDB; AAC78396.

XX Novel isolated nucleic acids comprising sequences encoding peptides  
PT useful for treating or diagnosing e.g. cancer -

XX Claim 11; Page 2302; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnary; immunomodulator;  
CC antidiabetic; antistatic; antirheumatic; antiallergic;  
CC antiinflammatory; antihypertensive; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;  
CC nootropic; vasotropic; antipruritic and antidiabetic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ

CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.  
XX

SQ Sequence 97 AA;  
Query Match 60.9%; Score 56; DB 21; Length 97;  
Best Local Similarity 66.7%; Pred. No. 1.1;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWGPNKDGKVG 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 13 PGPVGSKDGDPKG 27

RESULT 16

AAR79481  
ID AAR79481 standard; Protein: 54 AA.

AC AAR79481;

DT 19-JAN-1996 (first entry)

DE Mouse type II collagen.

XX Collagen; bone progenitor; gene transfer; gene therapy; osteoporosis;  
KW osteotomy; bone repair; osteotropic; plicha pastoris.

XX Mus sp.

OS WO9522611-A2.

PN 24-AUG-1995.

PD 21-FEB-1995; 95WO-US02251.

PR 30-SEP-1994; 94US-0316650.

PR 18-FEB-1994; 94US-0199780.

XX (UNMT) UNIV MICHIGAN.

PA WPI: 1995-302717/39.

DR Transferring nucleic acid into bone progenitor cell(s) - using a  
XX bone compatible matrix, for treatment of fracture(s) and  
PT osteoporosis.

XX Disclosure; Page 210; 317pp; English.

XX Human, rat and mouse collagen type II (given in AAR79479-81,  
CC respectively) can be used to stimulate bone progenitor cells as a  
CC means of treating bone-related diseases in association with an  
CC osteotropic gene.  
XX

SQ Sequence 54 AA;

Query Match 59.8%; Score 55; DB 16; Length 54;  
Best Local Similarity 66.7%; Pred. No. 0.87;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPWGPNKDGKVG 15  
| | | | | | | | | | | | | | | | | | | | | |  
Db 20 PGPVGSKDGDPKG 34

RESULT 17

AAR71703  
ID AAR71703 standard; Protein: 1418 AA.

AC AAR71703;



XX 17-OCT-1995 (first entry)  
 DT  
 XX  
 DE Collagen alpha 1 (II) chain precursor.  
 XX  
 XX  
 KM Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
 KM disorder; osteoporosis; metastatic progression; Paget's disease;  
 KM hyperthyroidism; bone; resorption; rheumatoid arthritis;  
 KM osteoarthritis; vasculitis syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9508115-A.  
 XX  
 PD 23-MAR-1995.  
 XX  
 PF 19-SEP-1994; 94WO-DK00348.  
 XX  
 PR 17-SEP-1993; 93DK-0001040.  
 XX  
 PA (OSTE-) OSTEOMETER AS.  
 XX  
 PI Bonde M, Qvist P;  
 XX  
 DR WPI; 1995-131456/17.  
 XX  
 PT Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptide(s) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 PT metabolism, e.g. osteoporosis.  
 XX  
 PS Disclosure (Appendix A); Page 53; 87pp; English.  
 XX  
 CC Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis, metastatic progression, Paget's disease,  
 CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degradation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vasculitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 XX  
 SQ Sequence 1418 AA;  
 SO  
 Query Match 59.8%; Score 55; DB 16; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 23;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 PAGPWGNGKDGKVG 15  
 DB 1096 PPGVGPSPGKDGANG 1110  
 ID AAY96124 standard; Peptide; 1418 AA.  
 AAY96124  
 AC AAY96124;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 DE Collagen type II alpha-1.  
 XX  
 KM Collagen type II; rheumatoid arthritis; osteoarthritis; assay;  
 KM diagnosis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6110689-A.  
 XX

PD 29-AUG-2000.  
 XX  
 PF 04-NOV-1997; 97US-0963825.  
 XX  
 PR 21-JAN-1994; 94US-0187319.  
 XX  
 PA (OSTE-) OSTEOMETER AS.  
 XX  
 PI Bonde M, Qvist P;  
 XX  
 DR WPI; 2000-586349/55.  
 XX  
 PT Assaying type I collagen fragments for diagnosing osteoporosis in  
 PT postmenopausal women, involves contacting body fluid with synthetic  
 PT collagen peptide and antibody and quantifying by competitive binding  
 PT assay -  
 XX  
 PS Disclosure; Column 37-46; 41pp; English.  
 XX  
 CC The present sequence is that of human type II collagen alpha-1.  
 CC The invention is based on the discovery of the presence of  
 CC particular collagen fragments in body fluids of patients compared  
 CC with those of healthy subjects. These fragments are generated  
 CC upon collagen degradation and are partly characterised by the  
 CC presence of potential sites for crosslinking. A method for  
 CC assaying collagen fragments in a body fluid sample is based on the  
 CC competitive binding to immunological binding partners of collagen  
 CC fragments in the sample and of synthetic peptides derived from  
 CC collagen and containing crosslinkable sites (see AAY96112-17). When  
 CC considering the degradation of type II collagen, the assay can be  
 CC used as a means of identifying the presence of rheumatoid arthritis  
 CC and osteoarthritis.  
 XX  
 SQ Sequence 1418 AA;  
 SO  
 Query Match 59.8%; Score 55; DB 21; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 23;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 PAGPWGNGKDGKVG 15  
 DB 1096 PPGVGPSPGKDGANG 1110  
 ID AAW61562 standard; protein; 1487 AA.  
 AAW61562  
 AC AAW61562;  
 XX  
 DT 02-NOV-1998 (first entry)  
 XX  
 DE Human type II collagen alpha-chain protein.  
 XX  
 KM Type II collagen alpha-chain; human; immunoassay; antibody; C-IIfree;  
 KM epitope; cartilage; enzyme-linked immunosorbent assay; ELISA; therapy;  
 KM diagnosis; arthritis; growth disorder; prognosis; drug screening;  
 KM anti-arthritis agent; matrix metalloproteinase inhibitor.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9835235-A1.  
 XX  
 PD 13-AUG-1998.  
 XX  
 PF 30-JAN-1998; 98WO-G800304.  
 XX  
 PR 06-FEB-1997; 97GB-0002252.  
 XX  
 PA (UYSH-) UNIV SHEFFIELD.  
 XX  
 PI Croucher LJ, Hollander AP;  
 XX

DR WPI: 1998-447376/38.  
 XX  
 XX Immunossay kit containing two antibodies recognising coupled  
 PT epitope(s) on collagen fragments - and new antibodies, for  
 PT diagnosing arthritis etc.; also prognosis and screening for  
 PT anti-arthritis agents or inhibitors of matrix metallo-protease  
 XX  
 PS Disclosure: Fig 2; 57pp; English.  
 XX  
 XX This sequence represents the human type II collagen alpha-chain which  
 CC is used in a method to produce an immunoassay kit comprising of two  
 CC antibodies (Ab1 and Ab2), mono- or poly-clonal, or their fragments,  
 CC that bind to two C-ii-free coupled epitopes (C-ii-free indicates any type  
 CC II collagen fragment that is released from degraded cartilage). The kits  
 CC are designed for sandwich immunoassays, specifically enzyme-linked  
 CC immunosorbent assay (ELISA), and C-ii-free is systemic (present in urine,  
 CC serum or synovial fluid). The kits are used for therapy, diagnosis (e.g.  
 CC routine screening for arthritis and other cartilage diseases, also to  
 CC diagnose growth disorders), prognosis (e.g. monitoring progression of  
 CC rheumatoid arthritis and osteoarthritis, or monitoring treatment with  
 CC growth hormone) and for drug screening (to identify, and assess efficacy  
 CC of, anti-arthritis agents and matrix metalloproteinase inhibitors).  
 CC C-ii-free, derived from the N-terminus of the alpha 1 chain, have  
 CC increased resistance to proteolysis, so can accumulate in vivo to a  
 CC concentration that allows accurate measurement by immunoassay.  
 CC  
 SQ Sequence 1487 AA;  
 XX  
 XX  
 Query Match 59.8%; Score 55; DB 19; Length 1487;  
 Best Local Similarity 66.7%; Pred. No. 24;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 PAGPMGPNCKDKGVG 15  
 DB 1165 PPGVPGSGCKDGANG 1179  
 XX  
 XX  
 RESULT 20  
 AAR71704  
 ID AAR71704 standard; protein; 1078 AA.  
 XX  
 AC AAR71704;  
 XX  
 DT 17-OCT-1995 (first entry)  
 XX  
 XX Collagen alpha 1 (III) chain precursor.  
 DE  
 XX  
 KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;  
 KW disorder; osteoporosis; metastatic progression; Paget's disease;  
 KW hyperthyroidism; bone; resorption; rheumatoid arthritis;  
 KW osteoarthritis; vasculitis syndrome.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09508115-A.  
 XX  
 PD 23-MAR-1995.  
 XX  
 PF 19-SEP-1994; 94WO-DK00348.  
 XX  
 PR 17-SEP-1993; 93DK-0001040.  
 XX  
 PA (OSTE-) OSTEOMETER AS.  
 XX  
 PI Bonde M, Qvist P;  
 XX  
 DR WPI: 1995-131456/17.  
 XX  
 XX Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptides) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 PT metabolism, e.g. osteoporosis.  
 CC  
 XX

PS Disclosure (Appendix A); Page 55; 87pp; English.  
 XX  
 XX Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis, metastatic progression, Paget's disease,  
 CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degradation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vasculitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 CC  
 SQ Sequence 1078 AA;  
 XX  
 XX  
 Query Match 58.7%; Score 54; DB 16; Length 1078;  
 Best Local Similarity 66.7%; Pred. No. 24;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 PAGPMGPNCKDKGVG 15  
 DB 994 PVGSGPCKDKMSG 1008  
 XX  
 XX  
 RESULT 21  
 AAY96125  
 ID AAY96125 standard; Peptide; 1078 AA.  
 XX  
 AC AAY96125;  
 XX  
 DT 19-DEC-2000 (first entry)  
 XX  
 XX Collagen type III alpha-1.  
 DE  
 XX  
 KM Collagen type III; vasculitis syndrome; assay; diagnosis.  
 KM  
 XX  
 OS Homo sapiens.  
 XX  
 PN US6110689-A.  
 XX  
 PD 29-AUG-2000.  
 XX  
 PF 04-NOV-1997; 97US-0963825.  
 XX  
 PR 21-JAN-1994; 94US-0187319.  
 XX  
 PA (OSTE-) OSTEOMETER AS.  
 XX  
 PI Bonde M, Qvist P;  
 XX  
 DR WPI: 2000-586349/55.  
 XX  
 XX Assaying type I collagen fragments for diagnosing osteoporosis in  
 PT postmenopausal woman. Involves contacting body fluid with synthetic  
 PT collagen peptide and antibody and quantifying by competitive binding  
 assay -  
 XX  
 PS Disclosure: Column 46-51; 41pp; English.  
 XX  
 XX The present sequence is that of human type III collagen alpha-1.  
 CC The invention is based on the discovery of the presence of  
 CC particular collagen fragments in body fluids of patients compared  
 CC with those of healthy subjects. These fragments are generated  
 CC upon collagen degradation and are partly characterised by the  
 CC presence of potential sites for crosslinking. A method for  
 CC assaying collagen fragments in a body fluid sample is based on the  
 CC competitive binding to immunological binding partners of collagen  
 CC fragments in the sample and of synthetic peptides derived from  
 CC collagen and containing crosslinkable sites (see AAY96118-21). When  
 CC considering the degradation of type III collagen, the assay can be  
 CC used as a means of identifying the presence of vasculitis syndrome.  
 CC  
 XX



```

FT Peptide 586..591 /label=GXGX'Y'_motif
FT Peptide 592..597 /label=GXGX'Y'_motif
FT Peptide 598..603 /label=GXGX'Y'_motif
FT Peptide 604..609 /label=GXGX'Y'_motif
FT Peptide 610..615 /label=GXGX'Y'_motif
FT Peptide 616..621 /label=GXGX'Y'_motif
FT Peptide 622..627 /label=GXGX'Y'_motif
FT Peptide 628..633 /label=GXGX'Y'_motif
FT Peptide 634..639 /label=GXGX'Y'_motif
FT Peptide 640..665 /label=GXGX'Y'_motif
FT Peptide 657..662 /label=GXGX'Y'_motif
FT Peptide 677..682 /label=GXGX'Y'_motif
FT Peptide 683..688 /label=GXGX'Y'_motif
FT Peptide 689..694 /label=GXGX'Y'_motif
FT Peptide 695..700 /label=GXGX'Y'_motif
FT Peptide 707..712 /label=GXGX'Y'_motif
FT Peptide 713..718 /label=GXGX'Y'_motif
FT Peptide 735..740 /label=GXGX'Y'_motif
FT Peptide 741..746 /label=GXGX'Y'_motif
FT Peptide 747..752 /label=GXGX'Y'_motif
FT Peptide 759..764 /label=GXGX'Y'_motif
FT Peptide 765..770 /label=GXGX'Y'_motif
FT Peptide 771..776 /label=GXGX'Y'_motif
FT Peptide 787..792 /label=GXGX'Y'_motif
FT Peptide 793..798 /label=GXGX'Y'_motif
FT Peptide 799..804 /label=GXGX'Y'_motif
FT Peptide 815..820 /label=GXGX'Y'_motif
FT Peptide 821..826 /label=GXGX'Y'_motif
FT Peptide 827..832 /label=GXGX'Y'_motif
FT Peptide 833..838 /label=GXGX'Y'_motif
FT Peptide 839..844 /label=GXGX'Y'_motif
FT Peptide 845..850 /label=GXGX'Y'_motif
FT Peptide 863..868 /label=GXGX'Y'_motif
FT Peptide 869..874 /label=GXGX'Y'_motif
FT Peptide 875..880 /label=GXGX'Y'_motif
FT Peptide 891..896 /label=GXGX'Y'_motif
FT Peptide 897..902 /label=GXGX'Y'_motif

```

```

FT Peptide /label=GXGX'Y'_motif
FT Peptide 903..908 /label=GXGX'Y'_motif
FT Peptide 911..916 /label=GXGX'Y'_motif
FT Peptide 917..922 /label=GXGX'Y'_motif
FT Peptide 928..933 /label=GXGX'Y'_motif
FT Peptide 934..939 /label=GXGX'Y'_motif
FT Peptide /label=GXGX'Y'_motif
FT Peptide 956..961 /label=GXGX'Y'_motif
FT Peptide 962..967 /label=GXGX'Y'_motif
FT Peptide 968..973 /label=GXGX'Y'_motif
FT Peptide 1126..1131 /label=GXGX'Y'_motif
FT Peptide 1145..1150 /label=GXGX'Y'_motif
FT Peptide 1193..1198 /label=GXGX'Y'_motif

```

```

XX US5643783-A.
XX
XX 01-JUL-1997.
XX
XX 01-DEC-1993; 93US-0159784.
XX
XX 01-DEC-1993; 93US-0159784.
XX
XX 01-DEC-1993; 93US-0159784.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Oh SP, Olsen BR.
XX
XX WPI: 1997-350247/32.
XX
XX N-PSDB: AAT84485.
XX
XX Nucleic acid encoding human alpha-1 collagen - for production of
XX recombinant alpha-1 collagen, for use in the treatment of cartilage
XX degeneration
XX
XX PS Disclosure; Fig 2; 35pp; English.

```

```

QY 1 PAGPWGPNKGDKVG 15
DB 341 PPGPGPGPNKGDKTRPG 355

Query Match 58.7%; Score 54; DB 18; Length 1288;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

RESULT 24
AAW92297
ID AAW92297 standard; peptide: 1288 AA.
XX
XX AAW92297;
AC
XX
XX 28-APR-1999 (first entry)
DT
XX
XX Mouse alpha-1 (XVIII) collagen chain common sequence MO18(common)28.
DE
XX
XX Human, type XVIII collagen; liver disease; cirrhosis; detection;
KW hepatocellular carcinoma; diagnosis.
XX
XX Mus sp.
XX
XX W09856399-A1.
XX
XX 17-DEC-1998.
XX
XX

```

PF 12-JUN-1998; 98WO-US12327.  
 XX  
 PR 12-JUN-1997; 97US-0049369.  
 XX  
 PA (FIF-) ACAD FINLAND.  
 PA (FIB-) FIBROGEN INC.  
 PA (INRM ) INST NAT SANTE & RECH MEDICALE.  
 XX  
 PI Clement B, Pihlajaniemi T, Rehn M;  
 XX WPI; 1999-070292/06.  
 DR  
 XX  
 PT diagnosis and monitoring of liver disease by measuring collagen type  
 PT XVIII levels - with elevated levels indicative of disease,  
 PT especially cirrhosis or hepatocellular carcinoma  
 XX  
 PS Example 6; Fig 8; 56pp; English.  
 XX  
 CC A method has been developed for the detecting liver disease. The method  
 CC comprises: (a) reacting a patient sample with antibodies (Ab) specific  
 CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen  
 CC complex (C) formed as indicator of the amount of Coll18 present; (c)  
 CC similar analysis of a non-diseased control; and (d) comparing the  
 CC amounts of Coll18 in the two samples to detect presence or progression of  
 CC disease. Elevated levels of Coll18 are: (1) indicative of disease,  
 CC specifically hepatocellular carcinoma (there is a relationship between  
 CC Coll18 mRNA levels and tumour size and necrosis, and survival times are  
 CC significantly higher in patients with higher Coll18 levels). The method  
 CC provides non-invasive, early and accurate diagnosis of liver disease.  
 CC The present sequence represents the sequence common to mouse alpha-1  
 CC (XVIII) collagen chain from the present invention.  
 XX  
 SO Sequence 1288 AA:  
 OY 1 PAGWPGNGDKGKVG 15  
 DB 341 PPGGPGPPGKDGTPG 355  
 XX  
 RESULT 25  
 ABB50291  
 ID ABB50291 standard; Protein: 1466 AA.  
 XX  
 AC ABB50291;  
 XX  
 DT 08-FEB-2002 (first entry)  
 XX  
 DE Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.  
 XX  
 KM Ovarian tumour marker gene; human; overexpression; upregulation;  
 KM epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
 KM identification; serous cystadenoma; borderline serous tumour;  
 KM serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KM mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KM undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
 KM adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KM immune response pathway; cell proliferation regulation; protein folding;  
 KM membrane localised; secreted; therapeutic target; cytostatic;  
 KM gene therapy; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175177-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 03-APR-2001; 2001WO-US10947.  
 XX

PR 03-APR-2000; 2000US-194336P.  
 XX  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX  
 PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 XX WPI; 2001-626450/72.  
 DR N-PSDB; ABA83117.  
 DR  
 XX  
 PT Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian  
 PT cancer treatment, by measuring expression level of ovarian tumor marker  
 PT gene -  
 XX  
 PS Claim 23; Page 114-117; 140pp; English.  
 XX  
 CC The invention relates to methods for diagnosing and prognosing ovarian  
 CC tumours in an individual via the detection and measurement of the  
 CC expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,  
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
 CC ABA83181 and ABA83183). The methods of the invention are useful for  
 CC detecting an ovarian tumour in a patient, for identifying an individual  
 CC at increased risk for developing ovarian cancer, in prognostic tests for  
 CC assessing the relative severity of ovarian cancer, in tests for  
 CC monitoring a patient in remission from ovarian cancer and in tests for  
 CC monitoring disease status in a patient being treated for ovarian cancer.  
 CC The methods can additionally be used to identify a particular tumour as  
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
 CC tumour. The ovarian tumour marker genes of the invention were identified  
 CC using SAGE (serial analysis of gene expression) and were found to be  
 CC overexpressed in a broad variety of ovarian epithelial tumour cells  
 CC relative to normal ovarian epithelial cells. The marker genes are  
 CC implicated in immune response pathways, in the regulation of cell  
 CC proliferation and in protein folding, and many of these are membrane-  
 CC localised or secreted. In addition to their use as diagnostic and  
 CC prognostic markers, the ovarian tumour marker genes or their encoded  
 CC proteins may be used as therapeutic targets for the treatment and  
 CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
 CC proteins encoded by ovarian tumour marker genes of the invention.  
 XX  
 SO Sequence 1466 AA:  
 OY 1 PAGWPGNGDKGKVG 15  
 DB 1141 PPGSGPPGKDGTSVG 1155  
 XX  
 RESULT 26  
 AAE02533  
 ID AAE02533 standard; Protein: 1466 AA.  
 XX  
 AC AAE02533;  
 XX  
 DT 10-AUG-2001 (first entry)  
 XX  
 DE Bovine alpha1(III) collagen #1.  
 XX  
 OS Bovine; alpha1(III) collagen; cytosolic; viral infection;  
 KM pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
 KM medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
 KM rheumatoid arthritis; beverage; photographic application.  
 XX  
 PN WO200134647-A2.  
 XX

```

XX 17-MAY-2001.
PD
XX
XX 10-NOV-2000; 2000MO-US30792.
PF
XX
XX 12-NOV-1999; 99US-0439058.
PR
XX 10-NOV-2000; 2000US-0439058.
PA
XX (FIBR-) FIBROGEN INC.
XX
XX Bell MP, Neff TB, Polarek JW, Seeley TW;
PI
XX WPI: 2001-335911/35.
DR
XX N-PSDB; AAD06574.
DR
XX
XX Novel isolated and purified bovine or porcine collagens and gelatins
PT useful in medical, pharmaceutical, food and cosmetic industries, as
PT vaccine, and for treating autoimmune disorders, infections and cancer
PT
PT
PS Example 2; Fig 4; 168pp; English.
PS
XX
XX The present sequence is bovine alpha1(III) collagen. The present
CC invention relates to recombinant synthesis of collagens and gelatins
CC derived from animals. Collagen is useful in medical, pharmaceutical,
CC food and cosmetic industries. Collagen is an important component of
CC arterial sealants, bone grafts, drug delivery system, dermal implants,
CC haemostats, and incontinence implants, and for treating autoimmune
CC disorders such as rheumatoid arthritis. Collagen is useful in food
CC products such as sausage casings, and in cosmetics or facial and skin
CC products such as moisturizers. Recombinant gelatin is useful in vaccine
CC formulations for treating viral infections, autoimmune diseases and
CC cancer. Gelatin is useful in the manufacture or as a component of
CC various pharmaceutical and medical devices and products, in food and
CC beverage industries, in hair care and skin care products, as a glue or
CC adhesive in various manufacturing processes, as a light-sensitive coating
CC in various electronic devices, as photoresist base in photolithographic
CC processes, in printing and photographic applications, in laboratory
CC application, and as a component in various gels used for biochemical and
CC electrophoretic analysis, including enzymographic gels.
CC
SQ Sequence 1466 AA;

```

```

Query Match 58.7%; Score 54; DB 22; Length 1466;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 1 PAGPWGPNKGKVG 15
   | | | | | | | | |
DB 1141 PVGSGGPPKDGASG 1155

```

```

RESULT 27
AAE02534
ID AAE02534 standard; Protein; 1466 AA.
XX
XX AAE02534;
AC
XX 10-AUG-2001 (first entry)
DT
XX
XX Bovine alpha1(III) collagen #2.
DE
XX
XX Bovine; alpha1(III) collagen; gelatin; cytosolic; viral infection;
KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
KW medical; arterial sealant; bone graft; dermal implant; hemostat; cancer;
KW rheumatoid arthritis; beverage; photographic application.
XX
XX Bos sp.
OS
XX
XX WO200134647-A2.
PN
XX
XX 17-MAY-2001.
PD
XX
XX

```

```

PF 10-NOV-2000; 2000MO-US30792.
XX
XX 12-NOV-1999; 99US-0439058.
PR
XX 10-NOV-2000; 2000US-0439058.
XX
XX (FIBR-) FIBROGEN INC.
XX
XX Bell MP, Neff TB, Polarek JW, Seeley TW;
PI
XX WPI: 2001-335911/35.
DR
XX
XX Novel isolated and purified bovine or porcine collagens and gelatins
PT useful in medical, pharmaceutical, food and cosmetic industries, as
PT vaccine, and for treating autoimmune disorders, infections and cancer
PT
PT
PS Example 2; Fig 6; 168pp; English.
PS
XX
XX The present sequence is bovine alpha1(III) collagen. The present
CC invention relates to recombinant synthesis of collagens and gelatins
CC derived from animals. Collagen is useful in medical, pharmaceutical,
CC food and cosmetic industries. Collagen is an important component of
CC arterial sealants, bone grafts, drug delivery system, dermal implants,
CC haemostats, and incontinence implants, and for treating autoimmune
CC disorders such as rheumatoid arthritis. Collagen is useful in food
CC products such as sausage casings, and in cosmetics or facial and skin
CC products such as moisturizers. Recombinant gelatin is useful in vaccine
CC formulations for treating viral infections, autoimmune diseases and
CC cancer. Gelatin is useful in the manufacture or as a component of
CC various pharmaceutical and medical devices and products, in food and
CC beverage industries, in hair care and skin care products, as a glue or
CC adhesive in various manufacturing processes, as a light-sensitive coating
CC in various electronic devices, as photoresist base in photolithographic
CC processes, in printing and photographic applications, in laboratory
CC application, and as a component in various gels used for biochemical and
CC electrophoretic analysis, including enzymographic gels.
CC
SQ Sequence 1466 AA;

```

```

Query Match 58.7%; Score 54; DB 22; Length 1466;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 1 PAGPWGPNKGKVG 15
   | | | | | | | | |
DB 1141 PVGSGGPPKDGASG 1155

```

```

RESULT 28
AAE02537
ID AAE02537 standard; Protein; 1466 AA.
XX
XX AAE02537;
AC
XX 10-AUG-2001 (first entry)
DT
XX
XX Porcine alpha1(III) collagen.
DE
XX
XX Porcine; alpha1(III) collagen; gelatin; cytosolic; viral infection;
KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
KW medical; arterial sealant; bone graft; dermal implant; hemostat; cancer;
KW rheumatoid arthritis; beverage; photographic application.
XX
XX Sus scrofa.
OS
XX
XX WO200134647-A2.
PN
XX
XX 17-MAY-2001.
PD
XX
XX 10-NOV-2000; 2000MO-US30792.
PF
XX
XX 12-NOV-1999; 99US-0439058.
PR
XX 10-NOV-2000; 2000US-0439058.

```

XX (FIBR-) FIBROGEN INC.  
 PA Bell MP, Neff TB, Polarek JW, Seeley TW;  
 XX WPI: 2001-335911/35.  
 DR N-PSDB; AAD06578.  
 XX  
 PT Novel isolated and purified bovine or porcine collagens and gelatins  
 PT useful in medical, pharmaceutical, food and cosmetic industries, as  
 PT vaccine, and for treating autoimmune disorders, infections and cancer  
 PT  
 XX  
 PS Example 5; Fig 12; 168pp; English.  
 CC The present sequence is porcine alpha(II) collagen. The present  
 CC invention relates to recombinant synthesis of collagens and gelatins  
 CC derived from animals. Collagen is useful in medical, pharmaceutical,  
 CC food and cosmetic industries. Collagen is an important component of  
 CC arterial sealants, bone grafts, drug delivery system, dermal implants,  
 CC haemostats, and incontinence implants, and for treating autoimmune  
 CC disorders such as rheumatoid arthritis. Collagen is useful in food  
 CC products such as sausage casings, and in cosmetics or facial and skin  
 CC formulations for treating viral infections, autoimmune diseases and  
 CC cancer. Gelatin is useful in the manufacture or as a component of  
 CC various pharmaceutical and medical devices and products, in food and  
 CC beverage industries, in hair care and skin care products, as a glue or  
 CC adhesive in various manufacturing processes, as a light-sensitive coating  
 CC in various electronic devices, as photoresist base in photolithographic  
 CC processes, in printing and photographic applications, in laboratory  
 CC application, and as a component in various gels used for biochemical and  
 CC electrophoretic analysis, including enzymographic gels.  
 XX  
 SQ Sequence 1466 AA;  
 Query Match 58.7%; Score 54; DB 22; Length 1466;  
 Best Local Similarity 66.7%; Pred. No. 33;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 PAGPMGPNCKDKVG 15  
 DB 1142 PVGPGSGPGKDGASG 1156  
 RESULT 29  
 ABG15191  
 ID ABG15191 standard; Protein: 1469 AA.  
 XX  
 AC ABG15191;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #15182.  
 XX  
 KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX

DR WPI: 2001-639362/73.  
 DR N-PSDB; AAS793378.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID No 45550; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1469 AA;  
 Query Match 58.7%; Score 54; DB 22; Length 1469;  
 Best Local Similarity 66.7%; Pred. No. 33;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 OY 1 PAGPMGPNCKDKVG 15  
 DB 1144 PVGPGSGPGKDGTSG 1158  
 RESULT 30  
 AAR53257  
 ID AAR53257 standard; Protein: 1838 AA.  
 XX  
 AC AAR53257;  
 XX  
 DT 12-JAN-1995 (first entry)  
 XX  
 DE Human collagen (Type V).  
 XX  
 KW Human collagen; alpha 1; V type collagen; placental mRNA.  
 XX  
 OS Homo sapiens.  
 OS  
 FH Key  
 FH Peptide  
 FT 1..38  
 FT Location/Qualifiers  
 FT Region  
 FT 444..538  
 FT /label= "signal\_peptide"  
 FT /note= "contains (Gly-X-Y) repeats"  
 FT Binding-site  
 FT 645..647  
 FT /label= RGD  
 FT /note= "cell adhesion motif"  
 FT Binding-site  
 FT 663..665  
 FT /label= RGD  
 FT /note= "cell adhesion motif"  
 FT Domain  
 FT 897..933  
 FT /label= heparin\_binding\_domain  
 FT 1573..1838  
 FT /label= C-terminal\_region  
 FT /note= "contains 8 Cys residues"

```

XX JP06105687-A.
PN
XX 19-APR-1994.
PD
XX 27-DEC-1991; 91JP-0358300.
PE
XX 27-DEC-1991; 91JP-0358300.
PR
XX 27-DEC-1991; 91JP-0358300.
PA
XX (TAKI ) TAKARA SHUZO CO LTD.
XX WPI, 1994-163129/20.
DR N-PSDB; AAQ64536.
XX
XX Human collagen V-type gene - is used for diagnosis of human
XX collagen V-type related diseases
XX
XX Claim 1; Page 6-14; 19pp; Japanese.
XX
XX This amino acid sequence of type V collagen contains several distinct
XX domains including a region comprising repeated (Gly-X-Y) motifs and a
XX central domain containing two copies of the Arg-Gly-Asp cell adhesion
XX motif. The cDNA sequence encoding type V collagen was isolated from a
XX human placental library and will be useful for diagnosis of diseases
XX related to type V collagen.
XX
XX Sequence 1838 AA;
SQ

```

```

Query Match 58.7%; Score 54; DB 15; Length 1838;
Best Local Similarity 66.7%; Pred. NO. 41;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 1 PAGPWGPNKDKVG 15
   ||||| | | | |
Db 1145 PAGPYGPPGDDGDKG 1159

```

Search completed: November 1, 2002, 12:52:35  
 Job time : 28.5 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:52:42 : Search time 10.5 Seconds  
(without alignments)  
34.894 Million cell updates/sec

Title: US-09-529-691a-3  
Perfect score: 92  
Sequence: 1 PAGWGPNGKDKVKG 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	68.5	534	US-09-029-348-5	Sequence 5, Appl1
2	63	68.5	535	US-09-029-348-1	Sequence 1, Appl1
3	63	68.5	537	US-09-029-348-4	Sequence 4, Appl1
4	63	68.5	1024	US-08-931-820-2	Sequence 2, Appl1
5	63	68.5	1366	US-08-963-825-19	Sequence 19, Appl1
6	63	68.5	1366	US-09-570-573-19	Sequence 19, Appl1
7	63	68.5	1366	US-09-548-608-19	Sequence 19, Appl1
8	59	64.1	1418	US-09-010-999-1	Sequence 1, Appl1
9	57	62.0	1442	US-08-316-650-12	Sequence 12, Appl1
10	57	62.0	1442	PCT-US95-02251-12	Sequence 12, Appl1
11	55	59.8	41	US-08-395-816-2	Sequence 2, Appl1
12	55	59.8	54	US-08-316-650-14	Sequence 14, Appl1
13	55	59.8	54	PCT-US95-02251-14	Sequence 14, Appl1
14	55	59.8	1060	US-08-931-820-3	Sequence 3, Appl1
15	55	59.8	1418	US-08-963-825-20	Sequence 20, Appl1
16	55	59.8	1418	US-09-570-573-20	Sequence 20, Appl1
17	55	59.8	1418	US-09-548-608-20	Sequence 20, Appl1
18	54	58.7	1057	US-08-931-820-4	Sequence 4, Appl1
19	54	58.7	1078	US-08-963-825-21	Sequence 21, Appl1
20	54	58.7	1078	US-09-570-573-21	Sequence 21, Appl1
21	54	58.7	1078	US-09-548-608-21	Sequence 21, Appl1
22	52	56.5	1057	US-08-931-820-1	Sequence 1, Appl1
23	52	56.5	1341	US-08-963-825-18	Sequence 18, Appl1
24	52	56.5	1341	US-09-570-573-18	Sequence 18, Appl1
25	52	56.5	1341	US-09-548-608-18	Sequence 18, Appl1
26	51	55.4	26	US-07-951-565-1	Sequence 1, Appl1
27	51	55.4	26	US-07-951-565-6	Sequence 6, Appl1

28	51	55.4	26	1	US-08-246-242-9	Sequence 9, Appl1
29	51	55.4	26	5	PCT-US96-00206-1	Sequence 1, Appl1
30	51	55.4	464	2	US-08-836-854-19	Sequence 19, Appl1
31	50	54.3	171	3	US-08-011-735-2	Sequence 2, Appl1
32	50	54.3	228	4	US-09-219-849-38	Sequence 38, Appl1
33	50	54.3	310	4	US-09-219-849-47	Sequence 47, Appl1
34	50	54.3	351	3	US-09-011-735-1	Sequence 1, Appl1
35	50	54.3	595	4	US-09-219-849-48	Sequence 48, Appl1
36	50	54.3	595	4	US-09-219-849-50	Sequence 50, Appl1
37	50	54.3	684	1	US-08-555-669-12	Sequence 12, Appl1
38	50	54.3	684	3	US-09-073-663-12	Sequence 12, Appl1
39	50	54.3	822	4	US-09-219-849-49	Sequence 49, Appl1
40	49	53.3	547	1	US-08-494-168-7	Sequence 7, Appl1
41	49	53.3	623	4	US-09-029-348-3	Sequence 3, Appl1
42	49	53.3	626	4	US-09-029-348-2	Sequence 2, Appl1
43	49	53.3	1694	1	US-08-494-168-2	Sequence 2, Appl1
44	48	52.2	160	1	US-08-479-233-11	Sequence 11, Appl1
45	48	52.2	160	5	PCT-US93-00643-11	Sequence 11, Appl1

## ALIGNMENTS

```

RESULT 1
US-09-029-348-5
: Sequence 5, Application US/09029348
: Patent No. 6171827
: GENERAL INFORMATION:
: APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
: TITLE OF INVENTION: NOVEL PROCOLLAGENS
: FILE REFERENCE: G087857PUS LISTING
: CURRENT APPLICATION NUMBER: US/09/029,348
: CURRENT FILING DATE: 1998-05-07
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 5
: LENGTH: 534
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
: OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-5

Query Match      68.5%  Score 63;  DB 4;  Length 534;
Best Local Similarity 73.3%  Pred. No. 0.18;
Matches 11;  Conservative 1;  Mismatches 3;  Indels 0;  Gaps 0;

Qy      1  PAGWGPNGKDKVKG 15
Db      223  PAGWGPNGKDKRTG 237

RESULT 2
US-09-029-348-1
: Sequence 1, Application US/09029348
: Patent No. 6171827
: GENERAL INFORMATION:
: APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
: TITLE OF INVENTION: NOVEL PROCOLLAGENS
: FILE REFERENCE: G087857PUS LISTING
: CURRENT APPLICATION NUMBER: US/09/029,348
: CURRENT FILING DATE: 1998-05-07
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 1
: LENGTH: 535
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
: OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-1

```

Query Match 68.5%; Score 63; DB 4; Length 535;  
Best Local Similarity 73.3%; Pred. No. 0.18;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
Db 224 PAGPSPGPKDGRGTG 238

RESULT 3  
US-09-029-348-4  
; Sequence 4, Application US/09029348  
; Patent No. 6171827  
; GENERAL INFORMATION:  
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER  
; TITLE OF INVENTION: NOVEL PROCOLLAGENS  
; FILE REFERENCE: D087857PUS LISTING  
; CURRENT APPLICATION NUMBER: US/09/029,348  
; CURRENT FILING DATE: 1998-05-07  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 537  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE  
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS  
US-09-029-348-4

Query Match 68.5%; Score 63; DB 4; Length 537;  
Best Local Similarity 73.3%; Pred. No. 0.18;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
Db 224 PAGPSPGPKDGRGTG 238

RESULT 4  
US-08-931-820-2  
; Sequence 2, Application US/08931820  
; Patent No. 6010863  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Assay for collagen degradation  
; NUMBER OF SEQUENCES: 4  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/931,820  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 96202596.1  
; FILING DATE:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1024 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Collagen type I  
US-08-931-820-2

Query Match 68.5%; Score 63; DB 3; Length 1024;  
Best Local Similarity 73.3%; Pred. No. 0.36;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
Db 976 PAGPSPGPKDGRGTG 990

RESULT 5  
US-08-963-825-19  
; Sequence 19, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: collagen alpha 2- type I  
US-08-963-825-19

Query Match 68.5%; Score 63; DB 3; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.49;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
Db 1055 PAGPSPGPKDGRGTG 1069

RESULT 6  
US-09-570-573-19  
; Sequence 19, Application US/09570573  
; Patent No. 6342361

GENERAL INFORMATION:  
APPLICANT: Ovis, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: collagen alpha 2- type I  
US-09-570-573-19  
Query Match 68.5%; Score 63; DB 4; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.49;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMGNGKDGKVG 15  
DB 1055 PAGPSGPKDGKRTG 1069  
RESULT 7  
US-09-548-608-19  
Sequence 19, Application US/09548608  
GENERAL INFORMATION:  
APPLICANT: Ovis, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York

COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/548,608  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1366 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: collagen alpha 2- type I  
US-09-548-608-19  
Query Match 68.5%; Score 63; DB 4; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.49;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 PAGPMGNGKDGKVG 15  
DB 1055 PAGPSGPKDGKRTG 1069  
RESULT 8  
US-09-010-999-1  
Sequence 1, Application US/09010999  
GENERAL INFORMATION:  
APPLICANT: Poole, Anthony R.  
APPLICANT: Hollander, Anthony P.  
APPLICANT: Billingham, R.C.  
TITLE OF INVENTION: IMMUNOSSAYS FOR THE MEASUREMENT OF  
TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,999  
FILING DATE: 22-JAN-1998  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/448,501  
FILING DATE: 17-JUL-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/984,123  
FILING DATE: 04-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 032931/0212  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1418 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Human Type II collagen  
US-09-010-999-1

Query Match 64.1%; Score 59; DB 4; Length 1418;  
Best Local Similarity 66.7%; Pred. NO. 1.8;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGPWPNGKDGKVG 15  
| | | | | : | | : |  
Db 1096 PGPWPSPGKDGAMG 1110

RESULT 9  
US-08-316-650-12  
Sequence 12, Application US/08316650  
Patent No. 5942496  
GENERAL INFORMATION:  
APPLICANT: Bonadio, Jeffrey  
APPLICANT: Roessler, Blake J.  
APPLICANT: Goldstein, Steven A.  
APPLICANT: Lin, Mushan  
TITLE OF INVENTION: METHODS AND COMPOSITIONS  
TITLE OF INVENTION: FOR STIMULATING BONE CELLS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 30-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC:008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1442 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-316-650-12

Query Match 62.0%; Score 57; DB 2; Length 1442;  
Best Local Similarity 73.3%; Pred. NO. 3.6;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWPNGKDGKVG 15  
| | | | | : | | : |  
Db 871 PGPWPNGKDGKVG 885

RESULT 10  
PCT-US95-02251-12  
Sequence 12, Application PC/TUS9502251  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE  
TITLE OF INVENTION: CELLS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version  
SOFTWARE: #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02251  
FILING DATE: CONCURRENTLY HERewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC009P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1442 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-02251-12

Query Match 62.0%; Score 57; DB 5; Length 1442;  
Best Local Similarity 73.3%; Pred. NO. 3.6;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWPNGKDGKVG 15  
| | | | | : | | : |  
Db 871 PGPWPNGKDGKVG 885



NAME: Parker, David L.  
REGISTRATION NUMBER: 32,165  
REFERENCE/DOCKET NUMBER: UMIC009P--  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US95-02251-14

Query Match 59.8%; Score 55; DB 5; Length 54;  
Best Local Similarity 66.7%; Pred. No. 0.22;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15  
DB 20 PPGPVGSGKDGANG 34

RESULT 14  
US-08-931-820-3  
Sequence 3, Application US/08931820  
Patent No. 6010863  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Assay for collagen degradation  
NUMBER OF SEQUENCES: 4  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/931,820  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: EP 96202596.1  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1060 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Collagen type II  
US-08-931-820-3

Query Match 59.8%; Score 55; DB 3; Length 1060;  
Best Local Similarity 66.7%; Pred. No. 4.9;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15  
DB 984 PPGPVGSGKDGANG 998

RESULT 15  
US-08-963-825-20  
Sequence 20, Application US/08963825  
Patent No. 6110689  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,825  
FILING DATE:  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1418 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN -ALPHA 1 (II)  
US-08-963-825-20

Query Match 59.8%; Score 55; DB 3; Length 1418;  
Best Local Similarity 66.7%; Pred. No. 6.7;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15  
DB 1096 PPGPVGSGKDGANG 1110

RESULT 16  
US-09-570-573-20  
Sequence 20, Application US/09570573  
Patent No. 6342361  
GENERAL INFORMATION:  
APPLICANT: Qvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022



RESULT 19  
US-08-963-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110689  
; GENERAL INFORMATION:  
; APPLICANT: Ovist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/187,319  
; FILING DATE: 21-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
US-08-963-825-21  
Query Match 58.7%; Score 54; DB 3; Length 1078;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
DB 994 PVGPGSPGKDGTS 1008  
RESULT 20  
US-09-570-573-21  
; Sequence 21, Application US/09570573  
; Patent No. 6342361  
; GENERAL INFORMATION:  
; APPLICANT: Ovist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/187,319  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
US-09-570-573-21  
Query Match 58.7%; Score 54; DB 4; Length 1078;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
DB 994 PVGPGSPGKDGTS 1008  
RESULT 21  
US-09-548-608-21  
; Sequence 21, Application US/09548608  
; Patent No. 6355442  
; GENERAL INFORMATION:  
; APPLICANT: Ovist, Per  
; APPLICANT: Bonde, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25



;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/548,608  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: 08/187,319  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: GOGORIS, Adda C  
;; REGISTRATION NUMBER: 29,714  
;; REFERENCE/DOCKET NUMBER: 4305/08701  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-527-7700  
;; TELEFAX: 212-753-6237  
;; INFORMATION FOR SEQ ID NO: 21:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1078 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; IMMEDIATE SOURCE:  
;; CLONE: COLLAGEN ALPHA 1 (III)  
US-09-548-608-21

Query Match 58.7%; Score 54; DB 4; Length 1078;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGNGKDGKVG 15  
Db 994 PVPSPGPRGDKGTSG 1008

RESULT 22  
US-08-931-820-1  
;; Sequence 1, Application US/08931820  
;; Patent No. 6010863  
;; GENERAL INFORMATION:  
;; APPLICANT:  
;; TITLE OF INVENTION: Assay for collagen degradation  
;; NUMBER OF SEQUENCES: 4  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/931,820  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: EP 96202596.1  
;; FILING DATE:  
;; INFORMATION FOR SEQ ID NO: 1:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1057 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; HYPOTHEICAL: NO  
;; ANTI-SENSE: NO  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; TISSUE TYPE: Collagen type I  
US-08-931-820-1

Query Match 56.5%; Score 52; DB 3; Length 1057;  
Best Local Similarity 60.0%; Pred. No. 13;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGNGKDGKVG 15  
Db 925 PVPSPGPRGDKGTG 939

RESULT 23  
US-08-963-825-18  
;; Sequence 18, Application US/08963825  
;; Patent No. 6110689  
;; GENERAL INFORMATION:  
;; APPLICANT: Ovist, Per  
;; APPLICANT: Bonde, Martin  
;; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
;; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
;; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
;; NUMBER OF SEQUENCES: 21  
;; DISORDERS ASSOCIATED WITH THE METABOLISM OF  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Darby & Darby PC  
;; STREET: 805 Third Avenue  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10022  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/963,825  
;; FILING DATE:  
;; CLASSIFICATION: 436  
;; PRIORITY APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/187,319  
;; FILING DATE: 21-JAN-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: GOGORIS, Adda C  
;; REGISTRATION NUMBER: 29,714  
;; REFERENCE/DOCKET NUMBER: 4305/08701  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 212-527-7700  
;; TELEFAX: 212-753-6237  
;; INFORMATION FOR SEQ ID NO: 18:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1341 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; ORIGINAL SOURCE:  
;; ORGANISM: Homo sapiens  
;; IMMEDIATE SOURCE:  
;; CLONE: COLLAGEN ALPHA 1 (I)  
US-08-963-825-18

Query Match 56.5%; Score 52; DB 3; Length 1341;  
Best Local Similarity 60.0%; Pred. No. 17;  
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGNGKDGKVG 15  
Db 962 PVPSPGPRGDKGTG 976

RESULT 24  
US-09-570-573-18  
;; Sequence 18, Application US/09570573  
;; Patent No. 6342361  
;; GENERAL INFORMATION:  
;; APPLICANT: Ovist, Per  
;; APPLICANT: Bonde, Martin

```

1      TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the
2      TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
3      TITLE OF INVENTION: Method and use of the Method to Diagnose the Presence of
4      TITLE OF INVENTION: Disorders Associated with the Metabolism of
5      NUMBER OF SEQUENCES: 21
6      CORRESPONDENCE ADDRESS:
7      ADDRESSEE: Darby & Darby PC
8      STREET: 805 Third Avenue
9      CITY: New York
10     STATE: New York
11     COUNTRY: USA
12     ZIP: 10022
13
14     COMPUTER READABLE FORM:
15     MEDIUM TYPE: Floppy disk
16     COMPUTER: IBM PC compatible
17     OPERATING SYSTEM: PC-DOS/MS-DOS
18     SOFTWARE: Patentin Release #1.0, Version #1.25
19     CURRENT APPLICATION DATA:
20     APPLICATION NUMBER: US/09/570,573
21     FILING DATE:
22     CLASSIFICATION:
23     PRIOR APPLICATION DATA:
24     APPLICATION NUMBER: 08/187,319
25     FILING DATE:
26     ATTORNEY/AGENT INFORMATION:
27     NAME: Gogoris, Adda C
28     REGISTRATION NUMBER: 29,714
29     REFERENCE/DOCKET NUMBER: 4305/08701
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: 212-527-7700
32     TELEFAX: 212-753-6237
33     TELEX: 236687
34     INFORMATION FOR SEQ ID NO: 18:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 1341 amino acids
37     TYPE: amino acid
38     TOPOLOGY: linear
39     MOLECULE TYPE: protein
40     ORIGINAL SOURCE:
41     ORGANISM: Homo sapiens
42     IMMEDIATE SOURCE:
43     CLONE: COLLAGEN ALPHA 1 (I)
44     US-09-570-573-18
45
46     Query Match          56.5%; Score 52; DB 4; Length 1341;
47     Best Local Similarity 60.0%; Pred. No. 17;
48     Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
49
50     QY      1 PAGEWGNKGKDKVYG 15
51             ||||| | | | : |
52 Db      962 PAGQGRGDKRGTEG 976
53
54 RESULT 25
55 US-09-548-608-18
56 Sequence 18, Application us/09548608
57 Patent No. 6355442
58 GENERAL INFORMATION:
59 APPLICANT: Qvist, Per
60 APPLICANT: Bonde, Martin
61 TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the
62 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
63 TITLE OF INVENTION: Method and use of the Method to Diagnose the Presence of
64 TITLE OF INVENTION: Disorders Associated with the Metabolism of
65 NUMBER OF SEQUENCES: 21
66 CORRESPONDENCE ADDRESS:
67 ADDRESSEE: Darby & Darby PC
68 STREET: 805 Third Avenue
69 CITY: New York
70 STATE: New York
71 COUNTRY: USA
72 ZIP: 10022
73
74 COMPUTER READABLE FORM:

```

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: Patentin Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/09/548, 608
7 FILING DATE:
8 CLASSIFICATION:
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: 08/187, 319
11 FILING DATE:
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Gogoris, Adda C
14 REGISTRATION NUMBER: 29, 714
15 REFERENCE/DOCKET NUMBER: 4305/08701
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 212-527-7700
18 TELEFAX: 212-753-6237
19
20 TELEX: 236687
21 INFORMATION FOR SEQ ID NO: 18:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 1341 amino acids
24 TYPE: amino acid
25 TOPOLOGY: linear
26 MOLECULE TYPE: protein
27 ORIGINAL SOURCE:
28 ORGANISM: Homo sapiens
29 IMMEDIATE SOURCE:
30 CLONE: COLLAGEN ALPHA 1 (I)
31
32 US-09-548-608-18
33
34 Query Match 56.5%, Score 52; DB 4; Length 1341;
35 Best local similarity 60.0%; Pred. No. 17;
36 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0.
37
38 OY 1 PAGPWGNGKDGKV 15
39 ||||| | | | |
40
41 Db 962 PAGPGRDKGETG 976
42
43 RESULT 26
44 US-07-951-565-1
45 Sequence 1, Application US/07951565
46 Patent No. 5399347
47 GENERAL INFORMATION:
48 APPLICANT: Trentham, David E.
49 APPLICANT: Welner, Howard L.
50 TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with
51 TYPE OF INVENTION: Type II Collagen
52 NUMBER OF SEQUENCES: 6
53 CORRESPONDENCE ADDRESS:
54 ADDRESSEE: Dardy & Dardy
55 STREET: 805 Third Ave.
56 CITY: New York
57 STATE: New York
58 COUNTRY: U.S.A.
59 ZIP: 10022-7513
60
61 COMPUTER READABLE FORM:
62 MEDIUM TYPE: Floppy disk
63 COMPUTER: IBM PC compatible
64 OPERATING SYSTEM: PC-DOS/MS-DOS
65 SOFTWARE: Patentin Release #1.0, Version #1.25
66 CURRENT APPLICATION DATA:
67 APPLICATION NUMBER: US/07/951,565
68 FILING DATE: 19920925
69 CLASSIFICATION: 424
70 ATTORNEY/AGENT INFORMATION:
71 NAME: Gogoris, Adda C.
72 REGISTRATION NUMBER: 29, 714
73 REFERENCE/DOCKET NUMBER: 1010/07300
74 TELECOMMUNICATION INFORMATION:
75 TELEPHONE: (212)527-7700
76 TELEFAX: (212)753-6237

```

TELEX: 236687  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: Internal  
; ORIGINAL SOURCE: Internal  
; ORGANISM: Gallus domesticus  
; TISSUE TYPE: collagen  
US-07-951-565-1

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMGPKGDKVKG 15  
1 PPGPLGPKGOTGELG 15

RESULT 27  
US-07-951-565-6  
; Sequence 6, Application US/07951565  
; Patent No. 5399347  
; GENERAL INFORMATION:

APPLICANT: Trencam, David E.  
APPLICANT: Weiner, Howard L.  
TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with  
TITLE OF INVENTION: Type II Collagen  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby  
STREET: 805 Third Ave.  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10022-7513

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/951,565  
FILING DATE: 19920925  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C.  
REGISTRATION NUMBER: 29,714  
REFERENCE/DOCKET NUMBER: 1010/07300  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)527-7700  
TELEFAX: (212)753-6237  
TELEX: 236687

INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; FRAGMENT TYPE: Internal  
US-07-951-565-6

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 PAGPMGPKGDKVKG 15  
1 PPGPLGPKGOTGELG 15

DB 1 PPGPLGPKGOTGELG 15

RESULT 28  
US-08-246-242-9  
; Sequence 9, Application US/08246242  
; Patent No. 5675060  
; GENERAL INFORMATION:  
APPLICANT: Benolst, C.  
APPLICANT: Mathis, D.  
APPLICANT: Kouskoff, V.  
TITLE OF INVENTION: Transgenic Arthritic Mice  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246,242  
FILING DATE: 19-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldstein, Jorge A.  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 1383.0080000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 26 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-246-242-9

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMGPKGDKVKG 15  
1 PPGPLGPKGOTGELG 15

RESULT 29  
PCT-US96-00206-1  
; Sequence 1, Application PC/TUS9600206  
; GENERAL INFORMATION:  
APPLICANT: Immunologic Pharmaceutical Corporation  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
TITLE OF INVENTION: TREATING RHEUMATOID ARTHRITIS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lappin & Kusmer  
STREET: 200 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/00206  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Keirner, Ann-Louise  
REGISTRATION NUMBER: 33,523  
REFERENCE/DOCKET NUMBER: 1M2-014PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-466-6000  
TELEFAX: 617-466-6040  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus type II collagen  
PCT-US96-00206-1

Query Match 55.4%, Score 51; DB 5; Length 26;  
Best Local Similarity 53.3%, Pred. No. 0.37;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGWGPNGKDGKVG 15  
| | | | | : : : : :  
Db 1 PTGPGPKGQTELG 15

## RESULT 30

US-08-836-854-19

Sequence 19, Application US/08836854  
Patent No. 5824547

## GENERAL INFORMATION:

APPLICANT: HASHINO, Kimikazu

APPLICANT: MATSUSHITA, Hideyuki

APPLICANT: KATO, Ikunoshin

TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W. Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/836,854

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/02425

FILING DATE: 29-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 317721/1994

FILING DATE: 29-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Browdy, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: HASHINO-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 464 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-836-854-19

Query Match 55.4%, Score 51; DB 2; Length 464;  
Best Local Similarity 66.7%, Pred. No. 7.6;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGWGPNGKDGKVG 15  
| | | | | : : : : :  
Db 442 PKGPPGPKDGLPG 456

Search completed: November 1, 2002, 12:54:45  
Job time : 11.5 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:38:32 : Search time 13.5 Seconds  
(without alignments)  
106.766 Million cell updates/sec

Title: US-09-529-691A-3

Perfect score: 92  
Sequence: 1 PAGPWGPNKGDKVGV 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database:

PIR\_71:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69	75.0	460	2 T33110	hypothetical prote
2	66	71.7	291	2 T20942	hypothetical prote
3	63	68.5	1366	1 CGH02S	collagen alpha 2(I
4	60	65.2	1373	1 A43291	collagen alpha 2(I
5	60	65.2	2944	2 A54849	collagen alpha 1(V
6	57	62.0	365	2 S10847	collagen alpha 2(I
7	57	62.0	1051	2 A35763	collagen alpha 2 C
8	57	62.0	1419	2 A41182	collagen alpha 1(I
9	57	62.0	1487	2 B41182	collagen alpha 1(I
10	56	60.9	193	2 S07133	collagen alpha 1(I
11	56	60.9	303	2 T19289	hypothetical prote
12	56	60.9	636	2 S41067	collagen alpha 1(I
13	56	60.9	888	2 S28791	collagen alpha 1(X
14	56	60.9	1076	2 T30842	serine-repeat anti
15	56	60.9	1486	1 B40333	collagen alpha 1(I
16	56	60.9	1492	1 A40333	collagen alpha 1(I
17	56	60.9	1806	1 CGH01E	collagen alpha 1(X
18	55	59.8	172	2 D41132	collagen-related p
19	55	59.8	326	2 T16841	hypothetical prote
20	55	59.8	363	2 T16831	hypothetical prote
21	55	59.8	464	2 S59513	collagen II A1 pro
22	55	59.8	615	2 A05269	collagen alpha 1(I
23	55	59.8	964	1 CGCH2S	collagen alpha 1(I
24	55	59.8	1042	1 CGCH1S	collagen alpha 1(I
25	55	59.8	1418	2 T45467	collagen alpha 1(I
26	55	59.8	1464	1 S59856	collagen alpha 1(I
27	55	59.8	1487	1 CGH06C	collagen alpha 1(I
28	55	59.8	1546	1 CGH02E	collagen alpha 2(X
29	54	58.7	228	2 A44982	collagen UC0L1 - p

30	54	58.7	289	2 T20177	hypothetical prote
31	54	58.7	299	2 T20605	hypothetical prote
32	54	58.7	300	2 T19929	hypothetical prote
33	54	58.7	302	2 T15936	hypothetical prote
34	54	58.7	305	2 S44767	hypothetical prote
35	54	58.7	330	2 T26004	hypothetical prote
36	54	58.7	418	2 T15142	hypothetical prote
37	54	58.7	435	2 T15143	hypothetical prote
38	54	58.7	633	2 B40983	collagen alpha 1(X
39	54	58.7	1049	1 CGB07S	collagen alpha 1(I
40	54	58.7	1120	2 H88449	protein F54D8.1(I
41	54	58.7	1315	2 A56101	collagen alpha 1(X
42	54	58.7	1466	1 CGH07L	collagen alpha 1(I
43	54	58.7	1774	2 B56101	collagen alpha 1(X
44	54	58.7	1838	1 CGH01V	collagen alpha 1(V
45	54	58.7	1843	2 S18603	collagen alpha 1(V

#### ALIGNMENTS

##### RESULT 1

T33110

hypothetical protein C18H7.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T33110

R:Tim-Wollam, A.; Fronick, W.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans cosmid C18H7.

A:Reference number: 221284

A:Accession: T33110

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-460 <TIN>

A:Cross-references: EMBL:AF067607; PIDN:AAC17641.1; GSPDB:GN00022; CESP:C18H7.3

A:Experimental source: strain Bristol N2; clone C18H7

C:Genetics:

A:Gene: CESP:C18H7.3

A:Map position: 4

A:Introns: 84/1

C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 75.0% Score 69; DB 2; Length 460;  
Best Local Similarity 80.0%; Pred. No. 0.017;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAGPWGPNKGDKVGV 15

DB 237 PAGPWGPNKGDKVGV 251

##### RESULT 2

T20942

hypothetical protein F15A2.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T20942

R:Gregory, J.

submitted to the EMBL Data Library, March 1996

A:Reference number: 219349

A:Accession: T20942

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-291 <MIL>

A:Cross-references: EMBL:270207; PIDN:CAA94128.1; GSPDB:GN00028; CESP:F15A2.1

A:Experimental source: clone F15A2

C:Genetics:

A:Gene: CESP:F15A2.1

A:Map position: X

A:Introns: 50/3

C:Superfamily: unassigned collagens

Query Match 71.7%; Score 66; DB 2; Length 291;  
Best Local Similarity 80.0%; Pred. No. 0.029;  
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 PAGPMPGPNKDKRGV 15  
|||||  
Db 183 PAGPAGPECKDGPVG 197

RESULT 3  
CGHUS  
N:Alternat names: procollagen alpha 2(I) chain  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1989 #sequence revision 25-Aug-1995 #text change 21-Jul-2000  
C:Accession: A28500; S00824; S09176; I55311; A58111; A28472; A42165; A34405; A90567; I553005; A02865  
R:de Wel, W.; Bernard, M.; Benson-Chanda, V.; Chu, M.L.; Dickson, L.; Weil, D.; Ramirez, J. Biol. Chem. 262, 16032-16036, 1987  
A:Title: Organization of the human pro-alpha-2(I) collagen gene.  
A:Reference number: A28500; MUID:88058962  
A:Accession: A28500  
A:Molecule type: DNA; mRNA  
A:Residues: 1-248, 'N', 250-1366 <DEM>  
A:Cross-references: GB:J03464; NID:G179595; PIDN:AA59374.1; PID:G179596  
R:Kuivaniemi, H.; Tromp, G.; Chu, M.L.; Prockop, D.J.  
Biochem. J. 252, 633-640, 1988  
A:Title: Structure of a full-length cDNA clone for the prepro-alpha-2(I) chain of human  
A:Reference number: S00824; MUID:88339824  
A:Accession: S00824  
A:Molecule type: mRNA  
A:Residues: 1-275, 'A', 277-332, 'V', 334-337, 'A', 339-482, 'A', 484-548, 'D', 550-765 <KU11>  
A:Cross-references: EMBL:Y00724; NID:G30022; PIDN:CAA68709.1; PID:G30023  
R:Dickson, L.A.; de Wel, W.; di Liberto, M.; Weil, D.; Ramirez, F.  
Nucleic Acids Res. 13, 3427-3438, 1985  
A:Title: Analysis of the promoter region and the N-propeptide domain of the human proalpha-2(I) collagen gene.  
A:Reference number: S09176; MUID:85242047  
A:Accession: S09176  
A:Molecule type: DNA  
A:Residues: 1-23; 33-58, 'P', 60-93 <DIC>  
A:Cross-references: EMBL:X02488; NID:G30098; PIDN:CAA56320.1; PID:G30099  
R:Well, D.; D'Alessio, M.; Ramirez, F.; Eyre, D.R.  
J. Biol. Chem. 265, 16007-16011, 1990  
A:Title: Structural and functional characterization of a splicing mutation in the pro-alpha-2(I) collagen gene.  
A:Reference number: I55311; MUID:90368825  
A:Accession: I55311  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 76-93 <WEI1>  
A:Cross-references: GB:M35391; NID:G189684; PIDN:AAA60041.1; PID:G189685  
A:Accession: A58111  
A:Molecule type: protein  
A:Residues: 23-75, 94-96 <WEI2>  
A:Note: mutant sequence from a patient with Ehlers-Danlos syndrome type VII  
R:Wirtz, M.K.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, D.W.  
J. Biol. Chem. 262, 16376-16385, 1987  
A:Title: Ehlers-Danlos syndrome type VIIB. Deletion of 18 amino acids comprising the N-terminus of the pro-alpha-2(I) collagen gene.  
A:Reference number: A28472; MUID:88059013  
A:Accession: A28472  
A:Molecule type: protein  
A:Residues: 32-75, 94-111 <WIR>  
A:Note: mutant sequence of patient with Ehlers-Danlos syndrome type VIIIB  
R:Chiodo, A.A.; Hockey, A.; Cole, W.G.  
J. Biol. Chem. 267, 6361-6369, 1992  
A:Title: A base substitution at the splice acceptor site of intron 5 of the COL1A2 gene  
s-Danlos syndrome type VII.  
A:Reference number: A42165; MUID:92210617  
A:Accession: A42165  
A:Molecule type: mRNA  
A:Residues: 50-126 <CHI>  
A:Note: parts of this sequence were determined by protein sequencing; a mutant sequence  
R:Well, D.; D'Alessio, M.; Ramirez, F.; Steinmann, B.; Wirtz, M.K.; Glanville, R.W.; Hol  
J. Biol. Chem. 264, 16804-16809, 1989

A:Title: Temperature-dependent expression of a collagen splicing defect in the fibroblast  
A:Reference number: A34405; MUID:89380311  
A:Accession: A34405  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 58-108 <WEI3>  
A:Cross-references: GB:J05049  
A:Note: the accession cited by the authors is not found in GenBank  
A:Note: parts of this sequence were determined by protein sequencing; a mutant having  
R:Click, E.M.; Bornstein, P.  
Biochemistry 9, 4699-4706, 1970  
A:Title: Isolation and characterization of the cyanogen bromide peptides from the alp  
A:Reference number: A90567; MUID:71038625  
A:Accession: A90567  
A:Molecule type: protein  
A:Residues: 72, 81, 'B', 83-96, 417-447 <CLT>  
A:Note: The compositions of peptides CNBR1, CNBR0, and CNBR2 were determined; evidence  
R:Kuivaniemi, H.; Sabol, C.; Tromp, G.; Sippola-Thiele, M.; Prockop, D.J.  
J. Biol. Chem. 263, 11407-11413, 1988  
A:Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen that  
is asymptomatic mother.  
A:Reference number: I55264; MUID:88298792  
A:Accession: I55264  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA; mRNA  
A:Residues: 145-197 <KU12>  
A:Cross-references: GB:M21671; NID:G189521; PIDN:AAA59994.1; PID:G553606  
A:Note: single base mutation in intron leads to abnormal splicing of mRNA  
R:Chapman, S.D.; Shapiro, J.R.; McKinstry, M.B.; Stover, M.L.; Branson, P.; Rowe, D.W.  
J. Bone Miner. Res. 7, 793-805, 1992  
A:Title: Expression of mutant alpha 1(I)-procollagen in osteoblast and fibroblast cult  
A:Reference number: I55485; MUID:92351816  
A:Accession: I55485  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 163-181, 200-213 <CH2>  
A:Cross-references: GB:S41099; NID:G252702; PIDN:AA822761.1; PID:G252703  
A:Note: mutant sequence from a patient with osteogenesis imperfecta type IV  
R:Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
J. Biol. Chem. 245, 5042-5048, 1970  
A:Title: Comparative study of glycopeptides derived from selected vertebrate collagen  
A:Reference number: A92069; MUID:71001508  
A:Accession: A92069  
A:Molecule type: protein  
A:Residues: 175-180 <MOR>  
A:Experimental source: SKR  
A:Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine  
R:Flitzek, P.P.; Furtmayr, H.; Kuehn, K.  
Eur. J. Biochem. 47, 257-261, 1974  
A:Title: Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-  
A:Reference number: A91224; MUID:75008198  
A:Accession: A91224  
A:Molecule type: protein  
A:Residues: 418-447 <RTE>  
R:Tromp, G.; Prockop, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 85, 5254-5258, 1988  
A:Title: Single base mutation in the pro alpha 2(I) collagen gene that causes effice  
A:Reference number: I59125; MUID:88276936  
A:Accession: I59125  
A:Status: translation not shown; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 520-573 <TRO>  
A:Cross-references: GB:M21353; NID:G180881; PIDN:AAA52053.1; PID:G180882  
A:Note: single base mutation in intron leads to splicing out of exon 28  
R:Bernard, M.P.; Myers, J.C.; Chu, M.L.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.  
Biochemistry 22, 1139-1145, 1983  
A:Title: Structure of a cDNA for the proalpha-2 chain of human type I procollagen. Co  
A:Reference number: S09174; MUID:83178919  
A:Accession: S09174  
A:Molecule type: mRNA  
A:Residues: 623-742, 'A', 744-764, 'X', 766-827, 'A', 829-830, 'P', 832-836, 'P', 838-1097, 'L',  
A:Cross-references: GB:J00115; GB:V00503; NID:G30123; PIDN:CAA23761.1; PID:G825646

A:Experimental source: skin fibroblast cells  
 R:Corlino, A.; Zolnerzi, F.; Valli, M.; Pignatelli, P.F.; Cetta, G.; Brunelli, P.C.; Montes  
 Hum. Mol. Genet. 3, 2201-2206, 1994  
 A:Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c  
 A:Reference number: 154363; PMID:95187161  
 A:Accession: 168663  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 663-675, 'V', 677, 'P', 679-742, 'A', 744-746 <ROR>  
 A:Cross-references: GB:I47668; NID:q1009095; PIDN:AAB59577.1; PID:q1009096  
 R:Kyte, J.C.; Drenth, J.; Byers, P.H.; Kyte, D.R.  
 J. Biol. Chem. 267, 23108-23112, 1992  
 A:Title: Incorporation of type I collagen molecules that contain a mutant alpha 2(I) cha  
 A:Reference number: 155369; PMID:93054637  
 A:Accession: 155369  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 665-666, 'D', 666-670 <NIY>  
 A:Cross-references: GB:L00613; NID:q180888; PIDN:AAB59384.1; PID:q180889  
 A:Note: mutant sequence from a patient with osteogenesis imperfecta  
 R:Bateman, J.F.; Hannagan, M.; Chan, D.; Cole, W.G.  
 Biochem. J. 276, 765-770, 1991  
 A:Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitut  
 e method.  
 A:Reference number: A56799; PMID:91291136  
 A:Accession: A56799  
 A:Molecule type: mRNA  
 A:Residues: 672-675, 'V', 677, 'P', 679-681 <BAT>  
 A:Cross-references: GB:S39878; NID:q1679911; PIDN:AAB19314.1; PID:q232761  
 A:Note: sequence extracted from NCBI backbone (NCBI:39878, NCBI:P:39886)  
 A:Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors sug  
 gested sequence  
 R:Meekel, J.K.; Vuorio, T.; Vuorio, E.  
 Biochem. Biophys. Acta 1049, 171-176, 1990  
 A:Title: Growth-dependent modulation of type I collagen production and mRNA levels in cu  
 A:Reference number: S10768; PMID:90304220  
 A:Accession: S10768  
 A:Molecule type: mRNA  
 A:Residues: 960-1021, 'I', 1023-1188, 'D', 1190-1197, 'S', 1199-1356 <MAR>  
 A:Cross-references: EMBL:X55525; NID:q30101; PIDN:CAAA59142.1; PID:q30102  
 A:Experimental source: fibroblast cell culture  
 R:Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.  
 Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981  
 A:Title: Cloning a cDNA for the pro-alpha2 chain of human type I collagen.  
 A:Reference number: A18855; PMID:81273090  
 A:Accession: A18855  
 A:Molecule type: mRNA  
 A:Residues: 964-979, 'V', 981-1018, 'Q', 1020 <MYE>  
 A:Cross-references: GB:J00114; NID:q180393; PIDN:AAAS196.1; PID:q180394  
 A:Note: 1019-Leu was also found  
 R:Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.  
 J. Biol. Chem. 263, 7734-7740, 1988  
 A:Title: Arginine for glycine substitution in the triple-helical domain of the products  
 A:Reference number: 155285; PMID:88227975  
 A:Accession: 155285  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1090-1107 <WEN>  
 A:Cross-references: GB:M2816; NID:q179602; PIDN:AAAS1844.1; PID:q179603  
 A:Accession: 170059  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1090-1101, 'R', 1103-1107 <WEN>  
 A:Cross-references: GB:M2817; NID:q179606; PIDN:AAAS1846.1; PID:q179607  
 A:Note: mutant sequence from a patient with osteogenesis imperfecta type IV  
 R:Myers, J.C.; Dickson, L.A.; de Wet, W.J.; Bernard, M.P.; Chu, M.L.; di Liberto, M.; Pe  
 J. Biol. Chem. 258, 10128-10135, 1983  
 A:Title: Analysis of the 3' end of the human pro-alpha-2(I) collagen gene. Utilization c  
 A:Reference number: S09175; PMID:83290853  
 A:Accession: S09175  
 A:Molecule type: DNA

Query Match 68.5%; Score 63; DB 1; Length 1366;

Best Local Similarity 73.3%; Pred. No. 0.38;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPMPNGDKGVG 15  
 Db 1055 PAGPSPGAKDGRGTG 1069

RESULT 4

Collagen alpha 2(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 10-Sep-1999

C:Accession: A43291; A54328

R:Phillips, C.L.; Morgan, A.L.; Lever, L.W.; Wenstrup, R.J.

Genomics 13, 1345-1346, 1992

A:Title: Sequence analysis of a full-length cDNA for the murine pro alpha 2(I) collag

A:Reference number: A43291; PMID:92372043

A:Accession: A43291

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1373 <PHI>

A:Cross-references: GB:X58251; NID:q50488; PIDN:CAAA1205.1; PID:q50489

A:Note: sequence extracted from NCBI backbone (NCBI:P:112027)

R:Phillips, C.L.; Lever, L.W.; Plimell, S.R.; Quarles, L.D.; Wenstrup, R.J.

J. Invest. Dermatol. 97, 980-984, 1991

A:Title: Construction of a full-length murine Proalpha2(I) collagen cDNA by the polym

A:Reference number: A54328; PMID:92084969

A:Accession: A54328

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-110 <PH2>

C:Genetics: 1

A:Gene: COL1A2

C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo

C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

F:1145-1373/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 65.2%; Score 60; DB 1; Length 1373;  
 Best Local Similarity 73.3%; Pred. No. 1;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPMPNGDKGVG 15  
 Db 1061 PAGPSPGAKDGRGTG 1075

RESULT 5

collagen alpha 1(VII) chain precursor - human

N:Alternate names: procollagen alpha 1(VII) chain

C:Species: Homo sapiens (man)

C:Date: 04-Nov-1994 #sequence-revision 04-Nov-1994 #text-change 20-Sep-1999

C:Accession: A54849; PH0844; S16316; I56328; A50296; I84686

R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.

J. Biol. Chem. 269, 20256-20262, 1994

A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(I)

A:Reference number: A54849; PMID:94327588

A:Accession: A54849

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-2944 <CHR>

A:Cross-references: GB:I02870; NID:q987124; PIDN:AAAT5438.1; PID:q987125

R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imanura, S.

Biochem. Biophys. Res. Commun. 183, 958-963, 1992

A:Title: Molecular cloning and characterization of type VII collagen cDNA.

A:Reference number: PH0844; PMID:92231902

A:Accession: PH0844

A:Molecule type: mRNA

A:Residues: 'E'F, 340-475, 'R'ALSTASHSTLCWRATRWPCNQGSHWTRACPCRNPAASHRAARG', 524-528, 'A';  
 A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:q453698; PIDN:BA02853.1; PID:q4536  
 A:Experimental source: keratinocyte  
 A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr

R.Parente, M.G.; Chung, L.C.; Rymaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991  
A>Title: Human type VII collagen: DNA cloning and chromosomal mapping of the gene.  
A:Reference number: S16316; MUID:91334380  
A:Accession: S16316  
A:Molecule type: mRNA  
A:Residues: 815-892, 'E', 894-1439 <PAR>  
A:CROSS-references: GB:M5156; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915  
A:Experimental source: keratinocyte  
R.Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisaianh, P.S.; Cook, M.E.; Wright, J.;  
J. Invest. Dermatol. 99, 691-696, 1992  
A>Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion prot  
A:Reference number: I56328; MUID:93107742  
A:Accession: I56328  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
A:CROSS-references: GB:S51236; NID:g262308; PIDN:AA24657.1; PID:g262309  
R.Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Gianville, R.W.; Burgess, R.E.  
J. Biol. Chem. 264, 3822-3826, 1989  
A>Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
A:Reference number: A30296; MUID:89139437  
A:Accession: A30296  
A:Molecule type: protein  
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041, '  
A>Note: two reported peptides cannot be reliably located  
R.Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A>Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
A:Reference number: I48103; MUID:93271985  
A:Accession: I48103  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 2395-2871, 'S', 2873-2944 <RE2>  
A:CROSS-references: GB:I06862; NID:g388713; PIDN:AAA89196.1; PID:g388714  
R.Christiano, A.M.; Rymaenen, M.; Uitto, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A>Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs  
A:Reference number: A55255; MUID:94224777  
A:Contents: annotation  
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit c  
ed and subsequently O-glycosylated.  
C:Genetics:  
A:Gene: GDB:COL7A1, EBR1, EBD1, EB  
A:CROSS-references: GDB:128750; OMIM:120120  
A:Map position: 3p21.3-3p21.3  
A>Note: defects in this gene can result in dominant and recessive dystrophic epidermolys  
A>Note: there are 118 introns  
C:Complex: type VII collagen is probably a homotrimer  
C:Function:  
A:Description: structural component of extracellular polymer associated with anchoring f  
C:Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homology; f  
C:Keywords: coll; extracellular matrix; glycoprotein; hydroxyllysine; hydroxyproli  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F:236-201/Domain: von Willebrand factor type A repeat homology <VWA1>  
F:327-413/Domain: fibronectin type III repeat homology <FN3>  
F:414-502/Domain: fibronectin type III repeat homology <FN2>  
F:508-593/Domain: fibronectin type III repeat homology <FN1>  
F:598-683/Domain: fibronectin type III repeat homology <FN2>  
F:686-771/Domain: fibronectin type III repeat homology <FN3>  
F:776-862/Domain: fibronectin type III repeat homology <FN2>  
F:864-952/Domain: fibronectin type III repeat homology <FN3>  
F:954-1045/Domain: fibronectin type III repeat homology <FN3>  
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
F:1170-1172/Region: cell attachment (R-G-D) motif  
F:1189-1253/Region: cysteine/proline-rich  
F:1254-2783/Region: interrupted helical  
F:1334-1336/Region: cell attachment (R-G-D) motif  
F:2008-2010/Region: cell attachment (R-G-D) motif  
F:2553-2555/Region: cell attachment (R-G-D) motif

F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F:2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>  
F:337-786, 1109/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:2167, 2176, 2185, 2188, 2664, 2667, 2673/Modified site: 4-hydroxyproline (Pro) #status ex  
F:2625, 2631/Modified site: 5-hydroxyllysine (Lys) #status experimental  
F:2625, 2631/Binding site: carboxylate (Lys) (covalent) #status experimental  
F:2634, 2802, 2804/Disulfide Bonds: Interchain #status predicted

Query Match 65.2%; Score 60; DB 2; Length 2944;  
Best Local Similarity 66.7%; Pred. No. 2.2;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGMPCNGKDGKVG 15  
DB 1623 PGPVGPGRGRDEGV 1637

RESULT 6

S10847 collagen alpha 2(I) chain - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 21-Nov-1993 #sequence\_revision 26-Jul-1996 #text\_change 23-May-1997

C:Accession: S10847; S65690

R.Dickson, L.A.; Nimmiya, Y.; Bernard, M.P.; Pesciotta, D.M.; Parsons, J.; Green, G.

J. Biol. Chem. 256, 8407-8413, 1981

A>Title: The exon/intron structure of the 3'-region of the pro-alpha-2(I) collagen ge

A:Reference number: S10847; MUID:81264246

A:Accession: S10847

A:Molecule type: DNA

A:Residues: 1-365 <DIC>

A:CROSS-references: EMBL:J00811

A>Note: the authors translated the codon ATA for residue 207 as Asp, AGC for residue

A:Accession: S65690

A:Molecule type: protein

A:Residues: 121-123, 'X', 125-134, 'X', 136-137, 'X', 139, 'XXX', 143 <DIX>

C:Genetics:

A:Gene: COL1A2

A:Introns: 36/3; 54/3; 90/3; 173/1; 236/3; 317/3

C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo

C:Keywords: coll; extracellular matrix; glycoprotein; trimer; triple helix

F:136-365/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.0%; Score 57; DB 2; Length 365;  
Best Local Similarity 66.7%; Pred. No. 0.74;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGMPCNGKDGKVG 15  
DB 56 PGPVGPGRGRGRNG 70

RESULT 7

A35763 collagen alpha 2 chain - sea urchin (Paracentrotus lividus) (fragment)

C:Species: Paracentrotus lividus (common urchin)

C:Date: 12-Oct-1990 #sequence\_revision 12-Oct-1990 #text\_change 20-Sep-1999

C:Accession: A35763

R.D'Alessio, M.; Ramirez, F.; Suzuki, H.R.; Solursh, M.; Gambino, R.

J. Biol. Chem. 265, 7050-7054, 1990

A>Title: Cloning of a fibrillar collagen gene expressed in the mesenchymal cells of t

A:Reference number: A35763; MUID:90216744

A:Accession: A35763

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1051 <DAAS>

A:CROSS-references: GB:J05422; NID:g159961; PIDN:AA29440.1; PID:g159962

C:Superfamily: unassigned collagens; fibrillar collagen carboxyl-terminal homology

F:830-1051/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 62.0%; Score 57; DB 2; Length 1051;  
Best Local Similarity 66.7%; Pred. No. 2.2;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;



OY 1 PAGPMPNGKDGKVG 15  
 ||||| ||||| : : :  
 Db 485 PAGPMPNGKDGKVG 499

## RESULT 8

B41182

collagen alpha 1(I) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 13-Aug-1999

C:Accession: A41182; A44885

R:Metsegeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A:Reference number: A41182; MUID:91358489

A:Accession: A41182

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1419 &lt;MET&gt;

A:Cross-references: GB:M65161

R:Chen, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.

Development 111, 945-953, 1991

A:Title: Expression of the mouse alpha 1(I) collagen gene is not restricted to cartilag

A:Reference number: A44885; MUID:91347939

A:Accession: A44885

A:Molecule type: DNA

A:Residues: 1-28 &lt;CHP&gt;

A:Cross-references: GB:S63190; NID:9234368; PIDN:AAB19627.1; PID:9234369

A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime

F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology &lt;FCC&gt;

Query Match 62.0% Score 57; DB 2; Length 1419;  
 Best Local Similarity 73.3% Pred. No. 2.9;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| ||||| : : :  
 Db 848 PAGPMPNGKDGKVG 862

## RESULT 9

B41182

collagen alpha 1(I) chain precursor (long splice form) - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 16-Jul-1999

C:Accession: B41182

R:Metsegeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991

A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and

A:Reference number: A41182; MUID:91358489

A:Accession: B41182

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-1487 &lt;MET&gt;

A:Cross-references: GB:M65161

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trime

F:1259-1487/Domain: fibrillar collagen carboxyl-terminal homology &lt;FCC&gt;

Query Match 62.0% Score 57; DB 2; Length 1487;  
 Best Local Similarity 73.3% Pred. No. 3.1;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| ||||| : : :  
 Db 916 PAGPMPNGKDGKVG 930

## RESULT 10

S07133

collagen alpha 1(I) chain precursor - chicken (fragment)

C:Species: Gallus gallus (chicken)

C:Date: 13-Jan-1995 #sequence\_revision 17-Jul-1998 #text\_change 13-Aug-1999

C:Accession: S07133; S24447

R:Deak, F.; Argraves, W.S.; Kles, I.; Sparks, K.J.; Goetlueck, P.F.

Biochem. J. 229, 189-196, 1985

A:Title: Primary structure of the telopeptide and a portion of the helical domain of

A:Reference number: S07133; MUID:85306862

A:Accession: S07133

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-193 &lt;DEA&gt;

A:Cross-references: EMBL:X02663; NID:963318; PIDN:CA26499.1; PID:9755732

R:van der Rest, M.; Mayne, R.

J. Biol. Chem. 263, 1615-1618, 1988

A:Title: Type IX collagen proteoglycan from cartilage is covalently cross-linked to t

A:Reference number: S23613; MUID:88115274

A:Accession: S24447

A:Molecule type: protein

A:Residues: 124-125, 'X', 127-129, 'NO' &lt;RES&gt;

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match 60.9% Score 56; DB 2; Length 193;  
 Best Local Similarity 66.7% Pred. No. 0.55;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| ||||| : : :  
 Db 47 PAGPMPNGKDGKVG 61

## RESULT 11

T19289

hypothetical protein C15A11.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T19289

R:Gardner, A.

submitted to the EMBL data library, September 1996

A:Reference number: Z19103

A:Accession: T19289

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-303 &lt;WIL&gt;

A:Cross-references: EMBL:Z79694; PIDN:CA801959.1; GSPDB:GN00019; CESP:C15A11.1

C:Experimental source: clone C15A11

A:Gene: CESP:C15A11.1

A:Map position: 1

C:Superfamily: unassigned collagens

Query Match 60.9% Score 56; DB 2; Length 303;  
 Best Local Similarity 66.7% Pred. No. 0.86;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| ||||| : : :  
 Db 210 PAGPMPNGKDGKVG 224

## RESULT 12

S41067

collagen alpha 1(I) chain - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999

C:Accession: S41067; A29905; S31924

R:Glumoff, V.; Maekela, J.K.; Vuorio, E.

Biochim. Biophys. Acta 1217, 41-48, 1994

A:Title: Cloning of cDNA for rat pro alpha-1(I) collagen mRNA. Different expression

A:Reference number: S41067; MUID:94114571

A:Accession: S41067

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-636 <GLI2>  
 A:Cross-references: EMBL:X70369; NID:g57915; PIDN:CAA49832.1; PID:g57916  
 R:Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lytle, C.R.; Kohn, B.; Mohn, K.  
 DNA 7, 347-354, 1988  
 A:Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2(V) collagen mRNAs by estr  
 A:Reference number: A29905; MUID:88296083  
 A:Accession: A29905  
 A:Molecule type: mRNA  
 A:Residues: 308-482 <FRA>  
 A:Cross-references: GB:M21354; NID:g203500; PIDN:AAA40942.1; PID:g203501  
 R:Glimoff, V.; Maekela, J.K.; Vuorio, E.  
 submitted to the EMBL Data Library, February 1993  
 A:Reference number: S31924  
 A:Accession: S31924  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 2-636 <GLI2>  
 A:Cross-references: EMBL:X70369  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F:408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 60.9%; Score 56; DB 2; Length 636;  
 Best Local Similarity 66.7%; Pred. No. 1.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDKGVG 15  
 DB 313 PVPMPGPNKDKGVG 327

## RESULT 13

collagen alpha 1(XI) chain - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C:Accession: S28791  
 R:Nan, H.D.; Barendse, M.; Upholt, W.B.  
 J. Biol. Chem. 267, 22581-22586, 1992  
 A:Title: The chicken alpha1(XI) collagen gene is widely expressed in embryonic tissues.  
 A:Reference number: S28791; MUID:93054557  
 A:Accession: S28791  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-888 <NAB>  
 A:Cross-references: EMBL:M88593; NID:g211619; PIDN:AAA48707.1; PID:g211620  
 C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
 F:665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 60.9%; Score 56; DB 2; Length 888;  
 Best Local Similarity 66.7%; Pred. No. 2.6;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDKGVG 15  
 DB 308 PVPMPGPNKDKGVG 322

## RESULT 14

serine-repeat antigen 3 - Plasmodium vivax  
 C:Species: Plasmodium vivax  
 C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
 C:Accession: T30842  
 R:Kiefer, M.C.; Crawford, K.A.; Boley, L.J.; Landsberg, K.E.; Gibson, H.L.; Kaslow, D.C.  
 Mol. Biochem. Parasitol. 78, 55-65, 1996  
 A:Title: Identification and cloning of a locus of serine repeat antigen (sera)-related g  
 A:Reference number: 220898; MUID:96408670  
 A:Accession: T30842  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1076 <KIE>  
 A:Cross-references: EMBL:U51723; NID:g1381087; PID:g1381090; PIDN:AAB41487.1

C:Genetics:  
 A:introns: 12/1; 253/1; 302/1  
 A>Note: V-SERA 3  
 C:Superfamily: Plasmodium vivax serine-repeat antigen

Query Match 60.9%; Score 56; DB 2; Length 1076;  
 Best Local Similarity 66.7%; Pred. No. 3.1;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDKGVG 15  
 DB 843 PVPMPGPNKDKGVG 857

## RESULT 15

collagen alpha 1(II) chain precursor - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: B40333  
 R:Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.  
 J. Cell Biol. 115, 565-575, 1991  
 A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis  
 A:Reference number: A40333; MUID:92011898  
 A:Accession: B40333  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1486 <SUA>  
 A:Cross-references: GB:M63595  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F:37-96/Domain: von Willebrand factor type C repeat homology <WMC>  
 F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 60.9%; Score 56; DB 1; Length 1486;  
 Best Local Similarity 66.7%; Pred. No. 4.3;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDKGVG 15  
 DB 1167 PVPMPGPNKDKGVG 1181

## RESULT 16

collagen alpha 1'(II) chain precursor - African clawed frog  
 C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 16-Jul-1999  
 C:Accession: A40333  
 R:Su, M.W.; Suzuki, H.R.; Bleker, J.J.; Solursh, M.; Ramirez, F.  
 J. Cell Biol. 115, 565-575, 1991  
 A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis  
 A:Reference number: A40333; MUID:92011898  
 A:Accession: A40333  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-1492 <SUA>  
 A:Cross-references: GB:M63596  
 A>Note: this sequence is presented as substitutions relative to another sequence in a  
 es they replace; the appropriate interpretation of the sequence figure was reconstruc  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F:37-96/Domain: von Willebrand factor type C repeat homology <WMC>  
 F:1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 60.9%; Score 56; DB 2; Length 1492;  
 Best Local Similarity 66.7%; Pred. No. 4.3;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDKGVG 15  
 DB 1170 PVPMPGPNKDKGVG 1184

## RESULT 17

CGHUIE

collagen alpha 1(XI) chain precursor - human

N:Alternate names: procollagen alpha 1(XI) chain

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1990 #sequence\_revision 03-Oct-1995 #text\_change 08-May-1998

C:Accession: A35239; A31795

R:Yoshida, H.; Ramirez, F.

J. Biol. Chem. 265, 6423-6426, 1990

A:Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and express

A:Reference number: A35239; MUID:90202924

A:Accession: A35239

A:Molecule type: mRNA

A:Residues: 1-558 &lt;YOS&gt;

A:Cross-references: GB:J05407

R:Bernard, M.; Yoshida, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;

J. Biol. Chem. 263, 17159-17166, 1988

A:Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type X

cartilagenous tissue.

A:Reference number: A92689; MUID:89034222

A:Accession: A31795

A:Molecule type: DNA; mRNA

A:Residues: 538-1806 &lt;BER&gt;

A:Cross-references: GB:J04177

A:Note: parts of this sequence were determined by protein sequencing

C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

ed and subsequently O-glycosylated.

C:Genetics:

A:Gene: GDB:COL1A1; COL16

A:Cross-references: GDB:120595; OMIM:120280

A:Map position: 1p21-1p21

A:Introns: 561/3; 579/3; 615/3; 633/3; 648/3; 666/3; 681/3

A:Note: the list of introns is incomplete

C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha

3(XI) chain (see PIR:CGH06c), initially linked by disulfide bonds among their carboxyl-

termed with desmosome cross-links made from lysine and allysine residues

C:Function:

A:Description: structural component of extracellular fibrous polymer associated with cell

A:Note: may play a role in controlling the lateral growth of collagen XI fibrils

C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

C:Keywords: coll; coll; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin

F:1-36/Domain: signal sequence #status predicted &lt;SIG&gt;

F:35-260/Domain: PAR-1-like #status predicted &lt;PAR&gt;

F:37-511/Domain: amino-terminal propeptide #status predicted &lt;PRO&gt;

F:512-1565/Produce: collagen alpha 1(XI) chain #status predicted &lt;MAT&gt;

F:512-577/Region: amino-terminal nonhelical telopeptide

F:528-1542/Region: helical

F:1543-1565/Region: carboxyl-terminal nonhelical telopeptide

F:1566-1806/Domain: carboxyl-terminal propeptide #status predicted &lt;CTP&gt;

F:1583-1805/Domain: fibrillar collagen carboxyl-terminal homology &lt;RCC&gt;

F:61-243,182-236/Disulfide bonds: #status predicted

F:505/Modified site: allysine (lys) #status predicted

F:612,1452/Modified site: 5-hydroxylysine (lys) #status predicted

F:612,1452/Binding site: carboxydrate (lys) (covalent) #status predicted

Query Match 60.9%; Score 56; DB 1; Length 1806;

Best Local Similarity 66.7%; Pred. No. 5.2;

Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWPNGKDGKVG 15

Db 1226 PRGPGPGAGDGPQG 1240

## RESULT 18

D41132 collagen-related protein 4 - Hydra magnipapillata (fragment)

C:Species: Hydra magnipapillata

C:Date: 05-Jun-1992 #sequence\_revision 12-Jun-1992 #text\_change 01-Dec-2000

C:Accession: D41132; S21932

R:Kurtz, E.M.; Holstein, T.W.; Petri, B.M.; Engel, J.; David, C.N.

J. Cell Biol. 115, 1159-1169, 1991

A:Title: Mini-collagens in hydra nematocytes.

A:Reference number: A41132; MUID:92064646

A:Accession: D41132

A&gt;Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-172 &lt;KOR&gt;

A:Cross-references: EMBL:X61048; NID:99452; PIDN:CAA43382.1; PID:99453

A:Note: the authors translated the codon GGT for residue 142 as Pro, and TTA for res1

C:Superfamily: unassigned collagens

Query Match 59.8%; Score 55; DB 2; Length 172;

Best Local Similarity 66.7%; Pred. No. 0.68;

Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPWPNGKDGKVG 15

Db 74 PMGPWPNGYDGOOG 88

## RESULT 19

T16841 hypothetical protein T10E10.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Jun-2000

C:Accession: T16841

R:Geisler, C.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T10E10.

A:Reference number: Z18588

A:Accession: T16841

A&gt;Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-326 &lt;GEI&gt;

A:Cross-references: EMBL:U39644; NID:91049339; PID:91049341; PIDN:AAA80358.1; CESP:T1

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:T10E10.2

C:Superfamily: unassigned collagens

Query Match 59.8%; Score 55; DB 2; Length 326;

Best Local Similarity 66.7%; Pred. No. 1.3;

Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWPNGKDGKVG 15

Db 167 PAGPWPNGSGSDG 181

## RESULT 20

T16831 hypothetical protein T07H6.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000

C:Accession: T16831

R:Geisler, C.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid T07H6.

A:Reference number: Z18586

A:Accession: T16831

A&gt;Status: preliminary; translated from GR/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-363 &lt;GEI&gt;

A:Cross-references: EMBL:U53344; NID:91255886; PID:91255887; PIDN:AAA96223.1; GSPDB:G

A:Experimental source: strain Bristol N2; clone T07H6

A:Gene: CESP:T07H6.3

A:Map position: X

C:Superfamily: unassigned collagens

Query Match 59.8%; Score 55; DB 2; Length 363;

Best Local Similarity 66.7%; Pred. No. 1.4;

Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 229 PAGPMPNGKDGKVG 243

RESULT 21  
S59513  
collagen II A1 protein - zebra fish (fragment)  
C:Species: Brachydanio rerio (zebra fish)  
C>Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 13-Aug-1999  
C:Accession: S59513  
R:Yan, Y.; Hotta, K.; Riggleman, B.; Postlethwait, J.H.  
submitted to the EMBL Data Library, March 1995  
A:Description: Expression of a type II collagen gene in the zebrafish embryonic axis.  
A:Reference number: S59513  
A:Accession: S59513  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-464 <YAM>  
A:Cross-references: EMBL:U23822; NID:g773660; PIDN:AAA96815.1; PID:g773661  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology <FCC>  
F:236-464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 59.8%; Score 55; DB 2; Length 464;  
Best Local Similarity 66.7%; Pred. No. 1.9;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 142 PSGPVGPAGKDGKNG 156

RESULT 22  
A05269  
collagen alpha 1(II) chain precursor - chicken (fragments)  
C:Species: Gallus gallus (chicken)  
C>Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 16-Jul-1999  
C:Accession: A05269; A38035; A20855  
R:Yamada, Y.; Iliou, G.; Mudryj, M.; Obici, S.; de Crombrughe, B.  
Nature 310, 333-337, 1984  
A:Title: Conservation of the sites for one but not another class of exons in two chick  
A:Reference number: A05269; MUID:84270696  
A:Accession: A05269  
A:Molecule type: DNA  
A:Residues: 1-329 <YAM1>  
A:Note: the authors translated the codon AAA for residue 68 as Arg, AAC for residue 173  
R:Yamada, Y.; Kuehn, K.; de Crombrughe, B.  
Nucleic Acids Res. 11, 2733-2744, 1983  
A:Reference number: A38035; MUID:83220816  
A:Accession: A38035  
A:Molecule type: DNA  
A:Residues: 330-615 <YAM2>  
A:Note: the authors translated the codon GAT for residue 548 as Glu  
R:Yamada, Y.; Mudryj, M.; Sullivan, M.; de Crombrughe, B.  
J. Biol. Chem. 258, 2758-2761, 1983  
A:Title: Isolation and characterization of a genomic clone encoding chick alpha type II  
A:Reference number: A20855; MUID:83135706  
A:Accession: A20855  
A:Molecule type: DNA  
A:Residues: 222-233, 'R', 235-236, 'V', 238-239, 330-336, 'V', 338-351 <YAM>  
A:Note: the authors translated the codon CGT for residue 234 as Gly, GTT for residue 237  
C:Comment: Chicken collagen alpha 1(II) chain has about 50 exons. This sequence corres  
9, 9(240-275), 6(276-293), 5(294-329), 4(330-423), 3(424-484), 2(485-567), and 1(568-615)  
C:Genetics:  
A:Introns: 68/3; 119/3; 134/3; 152/3; 188/3; 221/3; 239/3; 275/3; 293/3; 423/3; 48  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C:Keywords: coiled coil; extracellular matrix; glycoprotein; tandem repeat; trimer; tri  
F:1-68/Region: amino-terminal propeptide nonhelical  
F:2-62/Domain: von Willebrand factor type C repeat homology <WVC>  
F:69-347/Domain: collagen alpha helical chain (fragments) #status predicted <CAH>  
F:348-615/Domain: carboxyl-terminal propeptide #status predicted <CPR>

F:387-615/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F:347,348/Disulfide bonds: interchain #status experimental  
F:516/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 59.8%; Score 55; DB 2; Length 615;  
Best Local Similarity 66.7%; Pred. No. 2.5;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 295 PSGPMPNGKDGKRG 309

RESULT 23  
CGCH25  
collagen alpha 2(I) chain precursor - chicken (fragments)  
C:Species: Gallus gallus (chicken)  
C>Date: 24-Apr-1984 #sequence\_revision 15-Aug-1997 #text\_change 21-Jul-2000  
C:Accession: I50206; I50207; S07354; S10848; S10480; S11146; I50628; I50170; I50625;  
R:Boedtker, H.; Finer, M.; Aho, S.  
Ann. N. Y. Acad. Sci. 460, 85-116, 1985  
A:Title: The structure of the chicken alpha 2 collagen gene.  
A:Reference number: I50206; MUID:86185168  
A:Accession: I50206  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-245 <BOE>  
A:Cross-references: GB:M25963; NID:g211581; PIDN:AAA69960.1; PID:g211605  
A:Accession: I50207  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 246-431 <BO2>  
A:Cross-references: GB:M25965; NID:g211583; PIDN:AAA69961.1; PID:g211606  
R:Aho, S.; Tate, V.; Boedtker, H.  
Nucleic Acids Res. 12, 6117-6125, 1984  
A:Title: Location of the 11 bp exon in the chicken pro alpha-2(I) collagen gene.  
A:Reference number: S07354; MUID:84297217  
A:Accession: S07354  
A:Molecule type: DNA  
A:Residues: 1-33 <AHO>  
A:Cross-references: EMBL:X00760; NID:g63266; PIDN:CAA25330.1; PID:g63267  
R:Tate, V.E.; Finer, M.H.; Boedtker, H.; Doty, P.  
Nucleic Acids Res. 11, 91-104, 1983  
A:Title: Chick pro-alpha-2 (I) collagen gene: exon location and coding potential for  
A:Reference number: S10480; MUID:83246518  
A:Accession: S10480  
A:Molecule type: mRNA  
A:Residues: 1-89 <TAT>  
A:Cross-references: EMBL:X02657; NID:g63314; PIDN:CAA26493.1; PID:g63315  
A:Accession: S10480  
A:Molecule type: DNA  
A:Residues: 17-73 <TAM>  
A:Note: the authors translated the codon CAG for residue 42 as Glu  
R:Voegel, G.; Ohkubo, H.; Sobel, M.E.; Yamada, Y.; Pastan, I.; de Crombrughe, B.  
Proc. Natl. Acad. Sci. U.S.A. 78, 5334-5338, 1981  
A:Title: Structure of the promoter for chicken alpha-2 type I collagen gene.  
A:Reference number: S11146; MUID:82060240  
A:Accession: S11146  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-15 <VOG>  
A:Cross-references: EMBL:J00821  
R:Yamada, Y.; Avedimento, V.E.; Mudryj, M.; Ohkubo, H.; Voegel, G.; Iranl, M.; Pasta  
Cell 22, 887-892, 1980  
A:Title: The collagen gene: evidence for its evolutionary assembly by amplification of  
A:Reference number: I50170; MUID:81112157  
A:Accession: I50628  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 126-161 <YAM>  
A:Cross-references: EMBL:V00400; NID:g63305; PID:g833611  
A:Accession: I50170  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 126-161 <YA2>  
 A:Cross-references: GB:J00828; NID:g211295; PIDN:AAA51612.1; PID:g211317  
 A:Accession: I50625  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 530-558 <YA3>  
 A:Cross-references: EMBL:V00396; NID:g63295; PID:g833609  
 A:Accession: I50626  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 572-601 <YA4>  
 A:Cross-references: EMBL:V00398; NID:g63299; PID:g833610  
 A:Accession: I50624  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 433-483 <YA5>  
 A:Cross-references: EMBL:V00394; NID:g63287; PID:g833608  
 R:Kang, A.H.; Gross, J.  
 Biochemistry 9, 796-804, 1970  
 A:Title: Amino acid sequence of cyanogen bromide peptides from the amino-terminal region  
 A:Reference number: A90568; MWID:7013186  
 A:Accession: A90568  
 A:Molecule type: protein  
 A:Residues: 'Z', 79-92 <KAN>  
 A:Experimental source: skin  
 R:Kang, A.H.; Igarashi, S.; Gross, J.  
 Biochemistry 8, 3200-3204, 1969  
 A:Title: Characterization of the cyanogen bromide peptides from the alpha2 chain of chick  
 A:Reference number: A90557; MWID:69285369  
 A:Accession: A90557  
 A:Molecule type: protein  
 A:Residues: 'Z', 79-95; 398-409, 'A', 411, 'V', 413-428 <KAN2>  
 A:Experimental source: skin  
 A:Note: The compositions of the six CNBr peptides were determined. CNBr0 is residues 93-  
 R:Highberger, J.H.; Kang, A.H.; Gross, J.  
 Biochemistry 10, 610-616, 1971  
 A:Title: Comparative studies on the amino acid sequence of the alpha2-CN2 peptides from  
 A:Reference number: A90358; MWID:7115216  
 A:Accession: B90358  
 A:Molecule type: protein  
 A:Residues: 399-409, 'A', 411, 'V', 413-428 <HIG>  
 A:Experimental source: skin  
 R:Lane, J.M.; Miller, E.J.  
 Biochemistry 8, 2134-2139, 1969  
 A:Title: Isolation and characterization of the peptides derived from the alpha2 chain of  
 A:Reference number: A90555; MWID:69206862  
 A:Accession: A90555  
 A:Molecule type: protein  
 A:Residues: 'Z', 79-95; 398-409, 'A', 411, 'V', 413-428 <LAN>  
 A:Experimental source: bone  
 A:Note: the compositions of the six CNBr peptides were determined  
 A:Note: the compositions of CNBr1, CNBr0, and CNBr2 are identical with those from skin  
 R:Igarashi, S.; Kang, A.H.; Gross, J.  
 Biochem. Biophys. Res. Commun. 38, 697-702, 1970  
 A:Title: Renaturation and ordering by electron microscopy of the cyanogen bromide peptide  
 A:Content: annotation: skin, order of CNBr peptides  
 R:Vust, J.; Lane, J.M.; Fietzek, P.; Miller, E.J.; Piez, K.A.  
 Biochem. Biophys. Res. Commun. 38, 703-708, 1970  
 A:Title: The order of the CNBr peptides from the alpha2 chain of collagen.  
 A:Reference number: A90169; MWID:70181852  
 A:Accession: A90169  
 A:Contents: annotation: bone, order of CNBr peptides  
 R:Wozney, J.; Hanahan, D.; Tate, V.; Boedtker, H.; Doty, P.  
 Nature 294, 129-135, 1981  
 A:Title: Structure of the pro alpha-2(I) collagen gene.  
 A:Reference number: S07327; MWID:82058081  
 A:Accession: S07327  
 A:Molecule type: DNA  
 A:Residues: 74-92; 93, 247-431; 432, 693-774 <MOZ>  
 A:Cross-references: EMBL:J00826  
 R:Fuller, F.; Boedtker, H.  
 Biochemistry 20, 996-1006, 1981

A:Title: Sequence determination and analysis of the 3' region of chicken pro-alpha 1  
 A:Reference number: I50623; MWID:81160715  
 A:Accession: I50623  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 506-830, 'T', 831-903, 'N', 904-964 <FUL>  
 A:Cross-references: EMBL:V00390; NID:g63248; PIDN:CAA23668.1; PID:g63249  
 R:Avadimento, E.V.; Vogell, G.; Yamada, Y.; Matzel, J.V.  
 Cell 21, 689-696, 1980  
 A:Title: Correlation between splicing sites within an intron and their sequence comp  
 A:Reference number: I50172; MWID:81064671  
 A:Accession: I50172  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 536-558 <AVV>  
 A:Cross-references: GB:J00861; NID:g211323; PIDN:AAA8637.1; PID:g211326  
 R:Lehrach, H.; Fischbach, A.M.; Hanahan, D.; Wozney, J.; Fuller, F.; Crkvenjakov, R.;  
 Proc. Natl. Acad. Sci. U.S.A. 75, 5417-5421, 1978  
 A:Title: Construction and characterization of a 2.5-kilobase procollagen clone.  
 A:Reference number: I50171; MWID:79074829  
 A:Accession: I50171  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 484-505 <LEH>  
 A:Cross-references: GB:J00837; NID:g4530617; PIDN:AAA51614.1; PID:g211320  
 A:Accession: I50171  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: protein  
 A:Residues: 24/1, 27/3, 33/3, 45/3, 73/3, 92/3, 107/3, 125/3, 143/3, 161/3, 179/3, 197  
 A:Note: the list of introns is incomplete  
 C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydro  
 F:1-22/Domain: signal sequence #status predicted <SIG>  
 F:23-77/Domain: amino-terminal propeptide #status predicted <PRO>  
 F:78-964/Product: collagen alpha 2(I) chain (fragments) #status predicted <MATN>  
 F:78-88/Region: amino-terminal nonhelical telopeptide  
 F:737-964/Domain: fibrillar collagen carboxyl-terminal homology (FCC)  
 F:78/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone ca  
 F:82/Modified site: allylsine (Lys) #status experimental  
 F:422,425/Modified site: 4-hydroxyproline (Pro) #status experimental  
 F:666/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 59.8%; Score 55; DB 1; Length 964;  
 Best local Similarity 66.7%; Pred. No. 3.9;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGWPGNGDKRGV 15  
 DB 657 PPGSPGPGKDRNG 671

RESULT 24  
 CGCH15  
 collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)  
 C:Species: Gallus gallus (chicken)  
 C:Date: 12-Aug-1981 #sequence\_revision 06-Jul-1982 #text\_change 31-Mar-2000  
 C:Accession: A90458; A90181; A02857  
 R:Highberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross,  
 Biochemistry 21, 2048-2055, 1982  
 A:Title: Amino acid sequence of chick skin collagen alpha1(I)-C88 and the complete pr  
 A:Reference number: A90458; MWID:82231995  
 A:Accession: A90458  
 A:Molecule type: protein  
 A:Residues: 1-1036 <HIG>  
 A:Experimental source: skin  
 A:Note: this is the latest in a series of papers from these workers elucidating the s  
 R:Eyre, D.R.; Glimcher, M.J.  
 Biochem. Biophys. Res. Commun. 48, 720-726, 1972  
 A:Title: Evidence for a previously undetected sequence at the carboxyterminus of the  
 A:Reference number: A90181; MWID:72243016  
 A:Accession: A90181  
 A:Molecule type: protein  
 A:Residues: 1037-1042 <EYR>  
 A:Experimental source: skin

A>Note: residues 1037-1042 above correspond to the carboxyl end of the protein  
 C:Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some  
 C:Comment: Most of the prolines at the third position of the tripeptide repeating unit  
 C:Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pc  
 C:Superfamily: collagen alpha 1(I) chain, fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coll; extracellular matrix; glycoprotein; pyroglyutamic acid; trimer;  
 F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 59.8%; Score 55; DB 1; Length 1042;  
 Best Local Similarity 66.7%; Pred. No. 4.2;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| |||||  
 Db 60 PAGPMPNGKDGDEAG 74

## RESULT 25

collagen alpha 1(II) chain precursor [imported] - horse

N:Alternate names: type II collagen

C:Species: Equus caballus (domestic horse)

C>Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000 #text\_change 04-Mar-2000

C:Accession: T45467

R:Richardson, D.W.; Dodge, G.R.

A:Submitted to the EMBL Data Library, June 1996

A:Description: Cloning of equine type II collagen and modulation of its expression in eg

A:Reference number: 222977

A:Accession: T45467

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1418 <RIC>

A:Cross-references: EMBL:U62528; PIDN:AA05773.1

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 59.8%; Score 55; DB 2; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 5.7;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| |||||  
 Db 1096 PGPVGPNGKDGANG 1110

## RESULT 26

S59856

collagen alpha 1(III) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Aug-1999

C:Accession: S59856; S62120; S16373

R:Toman, P.D.; de Crombrughe, B.

Gene 147, 161-168, 1994

A>Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA

A:Reference number: S59856; MUID:95011609

A:Accession: S59856

A:Molecule type: DNA

A:Residues: 1-1464 <TOM>

A:Cross-references: EMBL:X52046

R:Toman, D.

A:Submitted to the EMBL Data Library, November 1994

A:Reference number: S62120

A:Accession: S62120

A:Molecule type: DNA

A:Residues: 1-866, 'G', 868-1464 <TOA>

A:Cross-references: EMBL:X52046; NID:9575321; PIDN:CAA36279.1; PID:9575322

R:Meisaeranta, M.; Toman, D.; de Crombrughe, B.; Viorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A>Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A:Reference number: S16176; MUID:91274355

A:Accession: S16373

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1442-1464 <MET>

A:Cross-references: EMBL:X57983; NID:950476; PIDN:CAA1048.1; PID:950477  
 C:Genetics: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3;  
 58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 940/3; 97  
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C:Keywords: coll; extracellular matrix  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-154/Domain: propeptide #status predicted <PRO>  
 F:32-92/Domain: von Willebrand factor type C repeat homology <VWC>  
 F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>  
 F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 59.8%; Score 55; DB 2; Length 1464;  
 Best Local Similarity 66.7%; Pred. No. 5.9;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 ||||| |||||  
 Db 1140 PVGPMPNGKDGKTS 1154

## RESULT 27

CCHUC

collagen alpha 1(II) chain precursor [validated] - human

N:Alternate names: procollagen alpha 1(II) chain

N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collag

C:Species: Homo sapiens (man)

C>Date: 28-May-1986 #sequence\_revision 01-Sep-1995 #text\_change 08-Dec-2000

C:Accession: A38513; S06715; S24270; A24838; S06496; A35428; A30147; A33116; S64674;  
 7250; 137251; 137252; 137253; 137254; 155338; 155353; 161910

R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.

Genomics 8, 41-48, 1990

A>Title: The human type II procollagen gene: identification of an additional protein-

A:Reference number: A38513; MUID:91184811

A:Accession: A38513

A:Molecule type: DNA

A:Residues: 1-103 <RYA>

A:Cross-references: GB:M60299; NID:9180883; PIDN:AAA3873.1; PID:9180884

R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.

Nucleic Acids Res. 17, 9473, 1989

A>Title: Nucleotide sequence of the full length cDNA encoding for human type II proco

A:Reference number: S06715; MUID:90067946

A:Accession: S06715

A:Molecule type: mRNA

A:Residues: 1-28, 'R', 99-1487 <SU2>

A:Cross-references: EMBL:X16468; NID:929515; PIDN:CAA34488.1; PID:929516

A>Note: alternative splice form 1

R:Vikkula, M.; Meisaeranta, M.; Syvaenen, A.C.; Ala-kokko, L.; Viorio, E.; Peltonen,

Biochem. J. 285, 287-294, 1992

A>Title: Structural analysis of the regulatory elements of the type-II procollagen ge

A:Reference number: S24270; MUID:92344585

A:Accession: S24270

A:Status: translation not shown

A:Molecule type: DNA

A:Residues: 1-28 <VIK>

A:Cross-references: EMBL:X58709; GB:S40537; NID:935659

A>Note: this translation is not annotated in Genbank entry HSPROCOE1, release 111.0

R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.

Gene 44, 11-16, 1986

A>Title: Promoter region of the human pro-alpha-1(II)-collagen gene.

A:Reference number: A24828; MUID:87031574

A:Accession: A24828

A:Molecule type: DNA

A:Residues: 1-8, 'R', 10-28 <NUN>

A:Cross-references: NID:M25698; NID:9180872; PIDN:AAA52051.1; PID:9559237

R:Balduin, C.T.; Reginaldo, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.

Biochem. J. 262, 521-528, 1989

A>Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(I

A:Reference number: S06496; MUID:90026318

A:Accession: S06496

A:Molecule type: mRNA

A:Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834,

A:Cross-references: EMBL:X16711; NID:930040; PIDN:CAA34683.1; PID:930041

A>Note: alternative splice form 1  
R:Ryan, M.C.; Sandell, L.J.  
J. Biol. Chem. 265, 10334-10339, 1990  
A:Title: Differential expression of a cysteine-rich domain in the amino-terminal propept  
A:Reference number: A35428; MUID:90285153  
A:Accession: A35428  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 27-81, 'V', 83-103 <RYA2>  
A>Note: alternative splice form 2' splicing appears to be under developmental regulation  
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.  
Genomics 4, 438-441, 1989  
A:Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf  
A:Reference number: A30147; MUID:89233138  
A:Accession: A30147  
A:Molecule type: DNA  
A:Residues: 104-157, 'P', 159-236 <SUM>  
A:Cross-references: GB:03065; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M2168; GB  
R:Ala-Rokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990  
A:Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri  
A:Reference number: A94227; MUID:90370826  
A:Accession: A33116  
A:Molecule type: DNA  
A:Residues: 171-172, 'C', 174-175 <ALA>  
A>Note: mutant sequence from a family with family with primary generalized osteoarthritis  
R:Diab, M.; Wu, J.J.; Eyre, D.R.  
Biochem. J. 314, 327-333, 1996  
A:Title: Collagen type IX from human cartilage: a structural profile of intermolecular c  
A:Reference number: 564673; MUID:96195147  
A:Accession: 564674  
A:Molecule type: protein  
A:Residues: 188-189, 'X', 191-195, 1224-1230, 'X', 1232-1236 <DIX>  
R:Franc, S.; Martin, E.; Bouillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,  
Eur. J. Biochem. 234, 125-131, 1995  
A:Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cat  
A:Reference number: 563514; MUID:96096730  
A:Accession: 563514  
A:Molecule type: protein  
A:Residues: 243-261, 575-590, 756-763, 'X', 765-779 <PXA>  
R:Tiller, G.E.; Weis, M.A.; Polunio, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,  
Am. J. Hum. Genet. 56, 388-395, 1995  
A:Title: An RNA-splicing mutation (G-51VS20) in the type II collagen gene (COL2A1) in a  
A:Reference number: 138867; MUID:95150028  
A:Accession: 138867  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TIL>  
A:Cross-references: EMBL:U15195; NID:9557053; PIDN:AB60370.1; PID:9557054  
R:Ramirez, F.  
submitted to the EMBL Data Library, December 1988  
A:Reference number: 504892  
A:Accession: 504892  
A:Molecule type: mRNA  
A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>  
A:Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050  
R:Viikula, M.; Peltonen, L.  
FEBS Lett. 250, 171-174, 1989  
A:Title: Structural analyses of the polymorphic area in type II collagen gene.  
A:Reference number: 505000; MUID:89325561  
A:Accession: 505000  
A:Molecule type: DNA  
A:Residues: 630-640, 'A', 642-785 <VIK2>  
A:Cross-references: EMBL:X16158; NID:929951; PIDN:CAA34278.1; PID:91335018; PIDN:CAA3427  
PIDN:CAA34283.1; PID:91335023; PIDN:CAA34284.1; PID:91335024  
R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D  
J. Biol. Chem. 267, 22522-22526, 1992  
A:Title: An amino acid substitution (G1853-->Glu) in the collagen alpha 1(II) chain pro  
A:Reference number: A44309; MUID:93054548  
A:Accession: A44309  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: DNA, mRNA  
A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',

A:Cross-references: GB:L00977; NID:9180812; PIDN:AB23914.1; PID:9258774  
A>Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence  
A>Note: this translation is not annotated and this publication is not cited in Genban  
A>Note: mutant sequence associated with perinatal lethal hypochondrogenesis  
R:Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3869-3893, 1990  
A:Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an indivi  
A:Reference number: S16502; MUID:90251662  
A:Accession: S16502  
A:Molecule type: DNA  
A:Residues: 1164-1164, 'GPSKDKANGIPGP', 1185-1199 <TI2>  
A:Cross-references: EMBL:M37126; NID:9180808; PIDN:AA52037.1; PID:9180809  
A>Note: mutant sequence from a patient with spondyloepiphyseal dysplasia  
R:Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.  
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985  
A:Title: Identification and characterization of the human type II collagen gene (COL2  
A:Reference number: A02858; MUID:85190534  
A:Accession: A02858  
A:Molecule type: DNA  
A:Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>  
A:Cross-references: GB:J00116; NID:9180395; PIDN:AA5197.1; PID:9180396  
R:Elima, K.; Vuorio, T.; Vuorio, E.  
Nucleic Acids Res. 15, 9499-9504, 1987  
A:Title: Determination of the single polyadenylation site of the human pro-alpha-1(II  
A:Reference number: A27280; MUID:88067771  
A:Accession: A27280  
A:Molecule type: DNA, mRNA  
A:Residues: 1175-1487 <ELI>  
A:Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097  
A:Experimental source: fetal epiphyseal cartilage  
R:van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.  
Biochem. J. 237, 923-925, 1986  
A:Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.  
A:Reference number: A57033; MUID:87099927  
A:Accession: A57033  
A:Molecule type: protein  
A:Residues: 'XE', 1244-1246, 'N', 1248, 'X', 1250-1265, 1295-1305, 1395-1408 <VAN>  
A>Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal p  
R:Strom, C.M.; Upholt, W.B.  
Nucleic Acids Res. 12, 1025-1038, 1984  
A:Title: Isolation and characterization of genomic clones corresponding to the human  
A:Reference number: A21733; MUID:84118798  
A:Accession: A21733  
A:Molecule type: DNA  
A:Residues: 1245-1295 <STR1>  
A:Cross-references: EMBL:X00339; EMBL:X00296; NID:9394699; PIDN:CAA25092.1; PID:94378  
A:Accession: B21733  
A:Molecule type: DNA  
A:Residues: 894-909, 'PE', <STR2>  
A:Cross-references: GB:K0185; NID:930035; PIDN:CAA25082.1; PID:91335032  
R:Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.  
Biochemistry 24, 6343-6348, 1985  
A:Title: Isolation and partial characterization of genomic clones coding for a human  
gene.  
A:Reference number: A24561; MUID:86104139  
A:Accession: A24561  
A:Molecule type: DNA  
A:Residues: 1296-1358 <MUN2>  
A:Cross-references: GB:M12048; NID:9180017  
A>Note: this translation is not annotated in Genbank entry HUMCCT2A, release 111.0  
A>Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with  
R:Sangiorji, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Telpouras, P.; Ramir  
Nucleic Acids Res. 13, 2207-2225, 1985  
A:Title: Isolation and partial characterization of the entire human pro alpha 1(II) c  
A:Reference number: I37249; MUID:85215609  
A:Accession: S59491  
A:Molecule type: DNA  
A:Residues: 7-28, 'R', 99-114, 541-578, 786-802, 1055-1056, 'N', 1058-1068, 'T', 1070-1109, 120  
A:Accession: I84453  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 7-28 <SN2>  
A:Cross-references: GB:M23759; NID:9180845; EMBL:X03320; GB:M24938; NID:9310104



A:Note: the GenBank PID is based on an incorrect reading frame  
A:Accession: J17250  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 541-560 <SANS>  
A:Cross-references: EMBL:X02378; GB:M23870; NID:g30107; PIDN:CAA26227.1; PID:g929622107  
A:Accession: J17251

Query Match	59.8%;	Score 55;	DB 1;	Length 1487;
Best Local Similarity	66.7%;	Pred. NO. 6;		
Matches 10; Conservative	1;	Mismatches 4;	Indels 0;	Gaps 0

```

OY      1  PAGPWGPNKGDKVCG 15
          ||| |||: ||| |
Db      1165 PPGVPVGPSSGKDGGANG 1179

```

RESULT 28

collagen alpha 2(XI) chain precursor - human (fragment)  
N:Alternate names: procollagen alpha 2(XI) chain  
N:Contains: proline/arginine-rich protein (PARP)  
C:Species: Homo sapiens (man)  
C:Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C:Accession: S34790; A32645  
R:Zhidkova, N.I.; Brewton, R.G.; Mayne, R.  
FEBS Lett. 326, 25-28, 1993  
A:Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage  
A:In: *Journal of Molecular Biology*, 230, 1993, 1-10

E.426.1266/Binding site: carboxylate (Lys) (covalent) #status predicted  
E.927.933.1008.1017.1035.1038.1290.1266.1305.1317.1320/Modified site: 4-hydroxyproline  
F.929/Modified site: 4-hydroxyproline (Pro) #status atypical  
F.942.1023.1289/Modified site: 5-hydroxylysine (Lys) #status experimental  
E.942.1023.1289/Binding site: carboxylate (Lys) (covalent) #status experimental  
F.1427.1423.1450.1455/Disulfide bonds: interchain #status predicted  
F.1460/Binding site: carboxylate (Asn) (covalent) #status predicted

query Match	59.88;	Score 55;	DB 1;	Length 1546;
Best Local Similarity	66.78;	Pred. No. 6.3;		
Matches 10;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0;

```
QY      1  PAGPWGPNKGDKVG  15
          | | | | | | |
Db      1040 PRGPAGPNGADGPQG 1054
```

RESULT 29

collagen UC001 pig roundworm (fragment)  
C:Species: Ascaris suum (pig roundworm)  
C:Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 20-Sep-1999  
C:Accession: A44982  
R:Kingston, I.B.; Walnwright, S.M.; Cooper, D.  
Mol. Biochem. Parasitol. 37, 137-146, 1989  
A:Title: Comparison of collagen gene sequences in Ascaris suum and Caenorhabditis elegans  
A:Reference number: A44982; MUID:90136710  
A:Accession: A44982  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-228 <RIN>  
A:Cross-references: GB:004657; NID:g159656; PIDN:AAA29371.1; PID:g159657  
A:Superfamily: unassigned collagens

Query Match	58.7%	Score 54	DB 2	length 228
Best Local Similarity	66.7%	Pred. No. 1.3		
Matches 10; Conservative	1	Mismatches 4	Indels 0	Gaps 0

```
QY      1 P A G P W G P N G K D G K V G      15
          | | | | | | | | |
Db      41 P P G P P G P D G K D G P D G      55
```

RESULT 30

hypothetical protein C53B4.5 - *Caenorhabditis elegans*  
C:Species: *Caenorhabditis elegans*  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C:Accession: T20177  
R:Berks, M.  
submitted to the EMBL data library, December 1995  
A:Reference number: Z19233  
A:Accession: T20177  
A:Status: preliminary; translated from GB/EMBL/DBD1  
A:Molecule type: DNA  
A:Residues: 1-289 <WIL>  
A:Cross-references: EMBL:Z68215; PIDN:CAA92453.1; GSPDB:GN00022; CESP:C53B4.5  
A:Experimental source: clone C53B4  
C:Genetics:  
A:Gene: CESP:C53B4.5  
A:Map position: 4  
A:Superfamily: unassigned collagens

Query Match	58.7%	Score 54	DB 2	Length 289
Best Local Similarity	66.7%	Pred. No.	1.6	
Matches 10	Conservative 0	Mismatches 5	Indels 0	Gaps 0

```
QY      1  PACPWPNGKDGKVG  15
          |||||  |||  |
Db      178  PACPPGPGGKPPGAG  192
```



Fri Nov 1 13:09:33 2002

us-09-529-691a-3.rpr

Page 13

Job time : 15.5 secs

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:16:06 : Search time 7 seconds

(without alignments)  
82.970 Million cell updates/sec

Title: US-09-529-691A-3  
Sequence: 1 PACMGPNGKDKVG 15

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	66	71.7	CA21_BOVIN	P02465 Bos taurus
2	64	69.6	CA21_ONCMY	O93484 oncorhynch
3	63	68.5	CA21_CANFA	O46392 canis fam1
4	63	68.5	CA21_HUMAN	P08123 homo sapien
5	62	67.4	CA21_RABIT	O28668 oryctolagus
6	60	65.2	CA21_MOUSE	O01119 mus musculu
7	60	65.2	CA21_RAT	P02466 rattus norv
8	60	65.2	CA17_HUMAN	Q02388 homo sapien
9	57	62.0	CA12_MOUSE	P28481 mus musculu
10	56	60.9	CA12_CHICK	P02460 gallus gall
11	56	60.9	CA13_RAT	P13941 rattus norv
12	56	60.9	CA1B_MOUSE	O61245 mus musculu
13	56	60.9	CA1B_HUMAN	P12107 homo sapien
14	55	59.8	CA13_CHICK	P12105 gallus gall
15	55	59.8	CA12_CHICK	P02467 gallus gall
16	55	59.8	CA12_HUMAN	P02468 homo sapien
17	55	59.8	CA11_CHICK	P02457 gallus gall
18	55	59.8	CA13_MOUSE	P08121 mus musculu
19	55	59.8	CA2B_MOUSE	O64739 mus musculu
20	55	59.8	CA2B_HUMAN	P13942 homo sapien
21	54	58.7	YK61_CAEEL	P34340 caenorhabd1
22	54	58.7	CA13_BOVIN	P02458 homo sapien
23	54	58.7	CA21_RANCA	O42350 rana catesb
24	54	58.7	CA13_HUMAN	P02461 homo sapien
25	54	58.7	CA1H_MOUSE	P39061 mus musculu
26	54	58.7	CA1H_HUMAN	P20908 homo sapien
27	53	57.6	CC36_CAEEL	P34803 caenorhabd1
28	53	57.6	CC06_CAEEL	P18831 caenorhabd1
29	53	57.6	CC06_CAEEL	P50404 mus musculu
30	53	57.6	PSPD_MOUSE	P35248 rattus norv
31	53	57.6	PSPD_RAT	P02453 bos taurus
32	53	57.6	CA11_BOVIN	P05957 homo sapien
33	52	56.5	CC02_CAEEL	P17656 caenorhabd1

## ALIGNMENTS

34	52	56.5	323	1	Y034_CAEEL	O09455 caenorhabd1
35	52	56.5	547	1	CAFL_EPHMU	P18856 ephydactyl m
36	52	56.5	675	1	CA39_CHICK	P32017 gallus gall
37	52	56.5	1453	1	CA11_MOUSE	P11087 mus musculu
38	52	56.5	1450	1	CA11_CANFA	O9X937 canis fam1
39	52	56.5	1464	1	CA11_HUMAN	P02452 homo sapien
40	52	56.5	1670	1	CA34_HUMAN	O01955 homo sapien
41	51	55.4	298	1	CC34_CAEEL	P34687 caenorhabd1
42	51	55.4	482	1	CA1B_RAT	P20909 rattus norv
43	51	55.4	680	1	CA1A_MOUSE	O05306 mus musculu
44	51	55.4	911	1	CA1B_BOVIN	O28083 bos taurus
45	51	55.4	1763	1	CA24_ASCSU	P27393 ascaris suu

## RESULT 1

ID	CA21_BOVIN	STANDARD	PRT	1364 AA.
AC	P02465; 062649;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 2(I) chain precursor.			
GN	COL1A2.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Aorta.			
RX	MEDLINE=98290219; PubMed=9628255;			
RA	Shirel T., Hattori S., Sakaguchi M., Inouye S., Kimura A., Ebihara T.,			
RA	Irie S., Nagai Y., Hori H.;			
RT	"The complete cDNA coding sequence for the bovine proalpha2(I) chain			
RT	of type I procollagen.";			
RL	Matrix Biol. 17:85-88(1998).			
RN	[2]			
RP	SEQUENCE OF 80-98.			
RC	TISSUE=Skin;			
RX	MEDLINE=75036115; PubMed=4609475;			
RA	Fietzek P.P., Breitkreutz D., Kuehn K.;			
RT	"Amino acid sequence of the amino-terminal region of calf skin			
RT	collagen.";			
RL	Biochim. Biophys. Acta 365:305-310(1974).			
RN	[3]			
RP	SEQUENCE OF 95-415, AND REVISION.			
RC	TISSUE=Skin;			
RX	MEDLINE=76091874; PubMed=173531;			
RA	Fietzek P.P., Rexrodt F.W.;			
RT	"The covalent structure of collagen. The amino-acid sequence of			
RT	alpha2-CB4 from calf-skin collagen.";			
RL	Eur. J. Biochem. 59:113-118(1975).			
RN	[4]			
RP	SEQUENCE OF 416-445.			
RC	TISSUE=Skin;			
RX	MEDLINE=75008198; PubMed=4412529;			
RA	Fietzek P.P., Furtmayr H., Kuehn K.;			
RT	"Comparative sequence studies on alpha2-CB2 from calf, human, rabbit			
RT	and pig-skin collagen.";			
RL	Eur. J. Biochem. 47:257-261(1974).			
RN	[5]			
RP	SEQUENCE OF 446-481.			
RC	TISSUE=Skin;			
RX	MEDLINE=75059250; PubMed=4435743;			
RA	Fietzek P.P., Kuehn K.;			
RT	"The covalent structure of collagen: amino acid sequence of the N-			
RT	terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5			
RT	from calf skin collagen.";			
RL	Hope-Seyler's Z. Physiol. Chem. 355:647-650(1974).			

```

CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB008683; BAA25171.1; -.
DR PIR; A02866; CGB025.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD002078; F1b_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Collagen; Signal.
FT SIGNAL 1 24
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 80 1100 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1101 1364 CARBOXYL-TERMINAL PROPEPTIDE.
FT MOD_RES 80 80 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 84 84 CONVERTED TO AN ALDEHYDE GROUP THAT IS
FT INVOLVED IN CROSS-LINKING (PROBABLE).
FT MOD_RES 175 175 HYDROXYLATION (PROBABLE).
FT MOD_RES 196 196 HYDROXYLATION (PROBABLE).
FT MOD_RES 262 262 HYDROXYLATION (PROBABLE).
FT MOD_RES 307 307 HYDROXYLATION (PROBABLE).
FT MOD_RES 352 352 HYDROXYLATION (PROBABLE).
FT CONFLICT 157 157 V -> P (IN REF. 3).
FT CONFLICT 187 187 K -> T (IN REF. 3).
FT CONFLICT 211 211 T -> K (IN REF. 3).
FT CONFLICT 288 300 PGA -> AGP (IN REF. 3).
FT CONFLICT 423 424 AT -> TA (IN REF. 4).
FT SEQUENCE 1364 AA; 129064 MW; 5593F4D6B9ED119A CRC64;
SQ
Query Match 71.7%; Score 66; DB 1; Length 1364;
Best Local Similarity 73.3%; Pred. No. 0.13;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 PAGPWGPNKGDKGVG 15
DB 1053 PAGPSPGACKDGRIG 1067

```

```

RT "Complete primary structure of rainbow trout type I collagen
RT consisting of alpha1(I)alpha2(I)alpha3(I) heterotrimers.";
RL Eur. J. Biochem. 268:2817-2827(2001).
RN [2]
RP SEQUENCE OF 417-1356 FROM N.A.
RC TISSUE=Fibroblast;
RA Salto M., Kunisaki N., Hirono I., Aoki T., Ishida M., Urano N.,
RA Kimura S.;
RT "Partial characterization of cDNA clones encoding the three distinct
RT pro alpha chains of type I collagen from rainbow trout.";
RL Fisheries Sci. 64:780-786(1998).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- P1M: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB052837; BAB55663.1; -.
DR EMBL; AB008372; BAA33379.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; F1b_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD002078; F1b_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KM Glycoprotein; Collagen; Signal.
FT SIGNAL 1 24
FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT CHAIN 80 1096 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1097 1356 CARBOXYL-TERMINAL PROPEPTIDE
FT (BY SIMILARITY).
FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 1356 AA; 126985 MW; 7BB2F1F80BD10C93 CRC64;
SQ
Query Match 69.6%; Score 64; DB 1; Length 1356;
Best Local Similarity 73.3%; Pred. No. 0.24;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1 PAGPWGPNKGDKGVG 15
DB 1049 PAGPHGPCKDGRAG 1063

```

```

RESULT 2
CA21_ONCMY STANDARD: PRT; 1356 AA.
AC 093484;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8022;
RN [1]
RP SEQUENCE FROM N.A.
RA Salto M., Takenouchi Y., Kunisaki N., Kimura S.;

```

```

RESULT 3
CA21_CANFA STANDARD: PRT; 1366 AA.
AC 046392;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC Tissue=Skin;
RA Campbell B.G., Wootton J.A.M., McLeod J.N., Minor R.R.;
RA Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.

```



RP REVIEW ON OI VARIANTS.  
 RX MEDLINE-91374476; PubMed-1895312;  
 RA Byers P.H., Wallis G.A., Willing M.C.;  
 RT "Osteogenesis imperfecta: translation of mutation to phenotype.";  
 RL J. Med. Genet. 28:433-442(1991).  
 RN [13]  
 RP REVIEW ON OI VARIANTS.  
 RX MEDLINE-97169389; PubMed-9016532;  
 RA Dalgleish R.;  
 RT "The human type I collagen mutation database.";  
 RL Nucleic Acids Res. 25:181-187(1997).  
 RN [14]  
 RP VARIANT EDS-VII-A2.  
 RX MEDLINE-88059013; PubMed-3680255;  
 RA Wirtz M.K., Gianvile R.W., Steinmann B., Rao V.H., Hollister D.W.;  
 RT "Ehlers-Danlos syndrome type VII. Deletion of 18 amino acids  
 competing the N-telopeptide region of a pro-alpha 2(I) chain.";  
 RL J. Biol. Chem. 262:16376-16385(1987).  
 RN [15]  
 RP SEQUENCE OF 1090-1107 FROM N.A., AND VARIANT OI-IV ARG-1102.  
 RX MEDLINE-88227975; PubMed-2897363;  
 RA Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;  
 RT "Arginine for glycine substitution in the triple-helical domain of  
 the products of one alpha 2(I) collagen allele (COL1A2) produces the  
 osteogenesis imperfecta type IV phenotype.";  
 RL J. Biol. Chem. 263:7734-7740(1988).  
 RN [16]  
 RP VARIANT OI-II ASP-997.  
 RX MEDLINE-89123407; PubMed-2914942;  
 RA Baldwin C.T., Constantinou C., Dunars K.W., Prockop D.J.;  
 RT "A single base mutation that converts glycine 907 of the alpha 2(I)  
 chain of type I procollagen to aspartate in a lethal variant of  
 osteogenesis imperfecta. The single amino acid substitution near the  
 carboxyl terminus destabilizes the whole triple helix.";  
 RL J. Biol. Chem. 264:3002-3006(1989).  
 RN [17]  
 RP VARIANT OI-II SER-955.  
 RX MEDLINE-89380165; PubMed-2777764;  
 RA Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;  
 RT "Characterization of point mutations in the collagen COL1A1 and  
 COL1A2 genes causing lethal perinatal osteogenesis imperfecta.";  
 RL J. Biol. Chem. 264:15809-15812(1989).  
 RN [18]  
 RP VARIANT OI-II CYS-877.  
 RA Fertala A., Westerhausen A., Morris G.M., Rooney J.E., Prockop D.J.;  
 RT "Two cysteine substitutions in the type I procollagen genes (COL1A1  
 and COL1A2) that cause lethal osteogenesis imperfecta. The location  
 of glycine substitutions does not in any simple way predict their  
 effects on protein function or phenotype.";  
 RL Am. J. Hum. Genet. 47:A216-A216(1990).  
 RN [19]  
 RP VARIANT EDS-VII-A2.  
 RX MEDLINE-90368825; PubMed-2394758;  
 RA Weil D., D'Alessio M., Ramirez F., Eyre D.R.;  
 RT "Structural and functional characterization of a splicing mutation in  
 the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII  
 patient.";  
 RL J. Biol. Chem. 265:16007-16011(1990).  
 RN [20]  
 RP VARIANTS OI-IV VAL-676.  
 RX MEDLINE-91291136; PubMed-2064612;  
 RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;  
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
 valine substitution in osteogenesis imperfecta type IV. Detection of  
 the mutation and prenatal diagnosis by a chemical cleavage method.";  
 RL Biochem. J. 276:765-770(1991).  
 RN [21]  
 RP VARIANTS OI CYS-349 AND CYS-736.  
 RX MEDLINE-91115889; PubMed-1990009;  
 RA Wenstrup R.J., Shrago-Howe A.W., Lever L.W., Phillips C.L.,  
 RA Byers P.H., Cohn D.H.;  
 RT "The effects of different cysteine for glycine substitutions within  
 alpha 2(I) chains. Evidence of distinct structural domains within the

RT type I collagen triple helix.";  
 RL J. Biol. Chem. 266:2590-2594(1991).  
 RN [22]  
 RP VARIANT OI-II ARG-784.  
 RX MEDLINE-91340689; PubMed-1874719;  
 RA Tsuneyoshi T., Westerhausen A., Constantinou C.D., Prockop D.J.;  
 RT "Substitutions for glycine alpha 1-637 and glycine alpha 2-694 of  
 type I procollagen in lethal osteogenesis imperfecta. The  
 conformational strain on the triple helix introduced by a glycine  
 substitution can be transmitted along the helix.";  
 RL J. Biol. Chem. 266:15608-15613(1991).  
 RN [23]  
 RP VARIANT OI-IV SER-751.  
 RX MEDLINE-91271401; PubMed-2052622;  
 RA Spocila L.D., Constantinou C.D., Sereida L., Ganguly A., Riggs B.L.,  
 RA Prockop D.J.;  
 RT "Mutation in a gene for type I procollagen (COL1A2) in a woman with  
 postmenopausal osteoporosis: evidence for phenotypic and genotypic  
 overlap with mild osteogenesis imperfecta.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5423-5427(1991).  
 RN [24]  
 RP VARIANT OI-II ARG-547.  
 RX MEDLINE-93244832; PubMed-1284475;  
 RA Bateman J.F., Moeller I., Hannagan M., Chan D., Cole W.G.;  
 RT "Lethal perinatal osteogenesis imperfecta due to a type I collagen  
 alpha 2(I) Gly to Arg substitution detected by chemical cleavage of  
 an mRNA: cDNA sequence mismatch.";  
 RL Hum. Mutat. 1:55-62(1992).  
 RN [25]  
 RP VARIANT OI-II ASP-670.  
 RX MEDLINE-93054637; PubMed-1385413;  
 QY 1 PAFPMGPNKGRKGV 15  
 DB 1055 PAFPMGPNKGRKGV 1069  
 Query Match 68.5%; Score 63; DB 1; Length 1366;  
 Best Local Similarity 73.3%; Pred. No. 0.33;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 RESULT 5  
 C21\_RABIT  
 ID C21\_RABIT STANDARD; PRT; 526 AA.  
 AC 028668;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor (Fragment).  
 GN COL1A2.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-New Zealand white; TISSUE-Calvaria;  
 RA Inoue S., Okazaki T.;  
 RT "Alpha 2 type I collagen gene expression in the rabbit knee ligaments:  
 variations during the newborn development and in the adult age.";  
 RL Biomed. Res. 16:219-227(1995).  
 CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 (FIBRILLAR FORMING COLLAGEN).  
 CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 HYDROXYAPATITE.  
 CC -I- PTV: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 DR EMBL; DA9399; BAA08391.1; -  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 FM Collagen.  
 FT NON\_TER 1 1  
 FT CHAIN <1 262 COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 263 526 CARBOXYL-TERMINAL PROPEPTIDE  
 FT (BY SIMILARITY).  
 SQ SEQUENCE 526 AA; 53129 MW; 7CA8F0FA6953846D CRC64;  
 Query Match 67.4%; Score 62; DB 1; Length 526;  
 Best Local Similarity 73.3%; Pred. No. 0.19;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 PAGPMPGNGKDGKYG 15  
 Db 215 PAGPMPGNGKDGKRG 229  
 RESULT 6  
 CA21\_MOUSE STANDARD; PRT; 1372 AA.  
 AC 001149;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor.  
 OS COLI12 OR COLA2.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Calvaria;  
 RX MEDLINE=92372043; Pubmed=1505972;  
 RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;  
 RT "Sequence analysis of a full-length cDNA for the murine pro alpha  
 RT 2(I) collagen chain: comparison of the derived primary structure with  
 RT human pro alpha 2(I) collagen.";  
 RL Genomics 13:1345-1346(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 1-110 FROM N.A.  
 RC TISSUE-Calvaria;  
 RX MEDLINE=92084969; Pubmed=1748823;  
 RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,  
 RA Wenstrup R.J.;  
 RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by  
 RT the polymerase chain reaction.";  
 RL J. Invest. Dermatol. 97:980-984(1991).  
 RN [4]  
 RP SEQUENCE OF 1-23 FROM N.A.  
 RX MEDLINE=87289650; Pubmed=3039494;  
 RA Rosol P., de Crombrughe B.;  
 RT "Identification of a cell-specific transcriptional enhancer in the  
 RT first intron of the mouse alpha 2 (type I) collagen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:5590-5594(1987).  
 CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN

CC (FIBRILLAR FORMING COLLAGEN).  
 CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -I- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>  
 CC or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC -----  
 DR EMBL; X58251; CAA41205.1; -  
 DR EMBL; BC007158; AAH07158.1; -  
 DR EMBL; K01832; AAA37331.1; -  
 DR PIR; A43291; AA3291.  
 DR MGD; MGI:88468; COL1a2.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Prodom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 85  
 FT CHAIN 86 1108 COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 1109 1372 CARBOXYL-TERMINAL PROPEPTIDE  
 FT (BY SIMILARITY).  
 FT MOD\_RES 86 86 PYRROLIDONE CARBOXYLIC ACID (BY  
 FT SIMILARITY).  
 FT MOD\_RES 90 90 CONVERTED TO AN ALDEHYDE GROUP THAT IS  
 FT INVOLVED IN CROSS-LINKING  
 FT (BY SIMILARITY).  
 FT CARBOHYD 1273 1273 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 15 15 V -> A (IN REF. 4).  
 FT CONFLICT 1167 1167 R -> TP (IN REF. 1).  
 SQ SEQUENCE 1372 AA; 129557 MW; 0017DE5DC1452D1 CRC64;  
 Query Match 65.2%; Score 60; DB 1; Length 1372;  
 Best Local Similarity 73.3%; Pred. No. 0.84;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 PAGPMPGNGKDGKYG 15  
 Db 1061 PAGPMPGNGKDGKRG 1075  
 RESULT 7  
 CA21\_RAT STANDARD; PRT; 1372 AA.  
 AC P02466; Q9RIE8;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor.  
 GN COL1A2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Guenther D., Seibold S., Marx M.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 86-98.

RC TISSUE-Skin; PubMed=5337886;  
 RA MEDLINE=67162268; PubMed=5337886;  
 RA Kang A.H., Bornstein P., Piez K.A.;  
 RT "The amino acid sequence of peptides from the cross-linking region of  
 RT rat skin collagen.";  
 RU Biochemistry 6:788-795(1967).  
 RN [3]  
 RP SEQUENCE OF 99-102.  
 RC TISSUE-Skin;  
 RX MEDLINE=69206881; PubMed=5785232;  
 RA Fietzek P.P., Piez K.A.;  
 RT "Isolation and characterization of the cyanogen bromide peptides from  
 RT the alpha 2 chain of rat skin collagen.";  
 RU Biochemistry 8:2129-2133(1969).  
 RN [4]  
 RP SEQUENCE OF 102-144.  
 RC TISSUE-Skin;  
 RX MEDLINE=73049496; PubMed=4636752;  
 RA Fietzek P.P., Kell I., Kuehn K.;  
 RT "The covalent structure of collagen. Amino acid sequence of the N-  
 RT terminal region of alpha 2-CB4 from calf and rat skin collagen.";  
 RU FEBS Lett. 26:66-68(1972).  
 RN [5]  
 RP SEQUENCE OF 423-452.  
 RC TISSUE-Skin;  
 RX MEDLINE=7115216; PubMed=5544653;  
 RA Highberger J.H., Kang A.H., Gross J.;  
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2  
 RT peptides from chick and rat skin collagens.";  
 RU Biochemistry 10:610-616(1971).  
 RN [6]  
 RP SEQUENCE OF 453-501.  
 RC TISSUE-Skin;  
 RX MEDLINE=75059250; PubMed=4435743;  
 RA Fietzek P.P., Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 RT terminal region of alpha 2-CB3 from rat skin collagen and alpha 2-CB3.5  
 RT from calf skin collagen.";  
 RU Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).  
 RN [7]  
 RP SEQUENCE OF 791-836.  
 RC TISSUE-Skin;  
 RX MEDLINE=74055004; PubMed=4763308;  
 RA Fietzek P.P., Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 RT terminal region of alpha 2-CB5 from rat skin collagen.";  
 RU FEBS Lett. 36:289-291(1973).  
 RN [8]  
 RP ORDER OF CNBR PEPTIDES.  
 RX MEDLINE=70181852; PubMed=5443712; Miller E.J., Piez K.A.;  
 RA Vuust J., Lane J.M., Fietzek P.P.;  
 RT "The order of the CNBR peptides from the alpha 2 chain of collagen.";  
 RU Biochem. Biophys. Res. Commun. 38:703-708(1970).  
 CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -I- PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF121217; AAD41775.1; -  
 DR PIR, A02867; CGRT25.  
 DR InterPro, IPR000087; Collagen.

DR InterPro, IPR000885; Fib\_collagen\_C.  
 DR Pfam, PF01391; Collagen\_18.  
 DR Prodom, PD002078; Fib\_collagen\_C; 1.  
 DR SMART, SM00038; COLFI; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 24  
 FT PROPEP 25 85  
 FT CHAIN 86 1108  
 FT PROPEP 1109 1372  
 FT SITE 783 785  
 FT SITE 828 830  
 FT SITE 1011 1013  
 FT MOD\_RES 86 86  
 FT MOD\_RES 90 90  
 FT CARBOHYD 1273 1273  
 FT CONFLICT 132 132  
 FT CONFLICT 137 137  
 FT CONFLICT 145 422  
 FT CONFLICT 431 432  
 FT CONFLICT 494 494  
 FT CONFLICT 497 497  
 FT CONFLICT 502 790  
 FT CONFLICT 825 825  
 SQ SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72 CRC64;  
 Query Match 65.2%; Score 60; DB 1; Length 1372;  
 Best Local Similarity 73.3%; Fred. No. 0.84;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 OY 1 PAGPMPNGKDGKVG 15  
 DB 1061 PAGPSPGPKDGRSG 1075  
 ID CA17\_HUMAN STANDARD; PRT; 2944 AA.  
 AC 002388; O14054; O16507;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VII) chain precursor (long-chain collagen) (LC  
 DE collagen).  
 GN COL7A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94327588; PubMed=8051117;  
 RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;  
 RT "Cloning of human type VII collagen. Complete primary sequence of the  
 RT alpha 1(VII) chain and identification of intragenic polymorphisms";  
 RU J. Biol. Chem. 269:20256-20262(1994).  
 RN [2]  
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=9338437; PubMed=1307247;  
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,  
 RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E.,  
 RA Uitto J.;  
 RT "The large non-collagenous domain (NC-1) of type VII collagen is  
 RT amino-terminal and chimeric. Homology to cartilage matrix protein,  
 RT the type III domains of fibronectin and the A domains of von  
 RT Willebrand factor.";  
 RU Hum. Mol. Genet. 1:475-481(1992).  
 RN [3]  
 RP SEQUENCE OF 815-1439 FROM N.A.  
 RX MEDLINE=91334380; PubMed=1871109;



RA Parente M.G., Chung L.C., Rymaenen J., Woodley D.T., Wynn K.W.,  
 RA Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;  
 RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the  
 RT gene."; *Proc. Natl. Acad. Sci. U.S.A.* 88:6931-6935(1991).  
 RL  
 RN [14]  
 RP SEQUENCE OF 369-1255 FROM N.A.  
 RX MEDLINE-93107742; PubMed-1469284;  
 RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisyanyh P.S.,  
 RA Cook M.E., Wright J., Briggaman R.A., Hunt S.W. III;  
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain  
 RT adhesion proteins involved in tissue-specific organization of  
 RT extracellular matrix."; *J. Invest. Dermatol.* 99:691-696(1992).  
 RL [5]  
 RP SEQUENCE OF 340-675 FROM N.A.  
 RC TISSUE-Keratinocytes;  
 RX MEDLINE-92231902; PubMed-1567409;  
 RA Tanaka T., Takahashi K., Furukawa F., Imanura S.;  
 RT "Molecular cloning and characterization of type VII collagen cDNA."; *Biochem. Biophys. Res. Commun.* 183:958-963(1992).  
 RL [6]  
 RP SEQUENCE OF 2395-2944 FROM N.A.  
 RX MEDLINE-93271985; PubMed-8499916;  
 RA Greenspan D.S.;  
 RT "The carboxyl-terminal half of type VII collagen, including the non-  
 RT collagenous NC-2 domain and intron/exon organization of the  
 RT corresponding region of the COL7A1 gene."; *Hum. Mol. Genet.* 2:273-278(1993).  
 RL [7]  
 RP SEQUENCE OF 1-87 FROM N.A.  
 RC TISSUE-Placenta;  
 RX MEDLINE-94375010; PubMed-8088784;  
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,  
 RA Uitto J., Greenspan D.S.;  
 RT "Structural organization of the human type VII collagen gene (COL7A1),  
 RT composed of more exons than any previously characterized gene."; *Genomics* 21:169-179(1994).  
 RL [8]  
 RP REVIEW ON DBE VARIANTS.  
 RX MEDLINE-98041696; PubMed-9375848;  
 RA Jaerykallio A., Pulkkinen L., Uitto J.;  
 RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in  
 RT the type VII collagen gene (COL7A1)."; *Hum. Mutat.* 10:338-347(1997).  
 RL [9]  
 RP VARIANT RDEB LYS-2798.  
 RX MEDLINE-93291877; PubMed-8513326;  
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y.,  
 RA Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;  
 RT "A missense mutation in type VII collagen in two affected siblings  
 RT with recessive dystrophic epidermolysis bullosa."; *Nat. Genet.* 4:62-66(1993).  
 RL [10]  
 RP VARIANT DDEB SER-2040.  
 RX MEDLINE-94224777; PubMed-8170945;  
 RA Christiano A.M., Rymaenen M., Uitto J.;  
 RT "Dominant dystrophic epidermolysis bullosa: identification of a  
 RT Gly->Ser substitution in the triple-helical domain of type VII  
 RT proc. *Natl. Acad. Sci. U.S.A.* 91:3549-3553(1994).  
 RL [11]  
 RP VARIANT PEB-DDEB CYS-2623.  
 RX MEDLINE-96081220; PubMed-8541842;  
 RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;  
 RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and  
 RT identification of a glycine-to-cysteine substitution in the triple-  
 RT helical domain of type VII collagen."; *Hum. Mol. Genet.* 4:1579-1583(1995).  
 RL [12]  
 RP VARIANT DDEB ARG-2043.  
 RX MEDLINE-95164985; PubMed-7861014;  
 RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C.,  
 RA Cavalleri R., Uitto J.;  
 RT "A glycine-to-arginine substitution in the triple-helical domain of  
 RT type VII collagen in a family with dominant dystrophic epidermolysis  
 RT bullosa."; *Invest. Dermatol.* 104:438-440(1995).  
 RL [13]  
 RP VARIANTS RDEB AND DDEB.  
 RX MEDLINE-96202218; PubMed-8644729;  
 RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;  
 RT "Glycine substitutions in the triple-helical region of type VII  
 RT collagen result in a spectrum of dystrophic epidermolysis bullosa  
 RT phenotypes and patterns of inheritance."; *Am. J. Hum. Genet.* 58:671-681(1996).  
 RL [14]  
 RP VARIANT RDEB ARG-2575.  
 RX MEDLINE-96154068; PubMed-8592061;  
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;  
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:  
 RT genotype/phenotype correlation in a case of moderate clinical  
 RT severity."; *J. Invest. Dermatol.* 106:119-124(1996).  
 RL [15]  
 RP VARIANT RDEB ARG-1782.  
 RX MEDLINE-96183562; PubMed-8618018;  
 RA Christiano A.M., McGrath J.A., Uitto J.;  
 RT "Influence of the second COL7A1 mutation in determining the  
 RT phenotypic severity of recessive dystrophic epidermolysis bullosa."; *J. Invest. Dermatol.* 106:766-770(1996).  
 RL [16]  
 RP VARIANT RDEB ASP-2073.  
 RX MEDLINE-96310789; PubMed-8757758;  
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,  
 RA Uitto J., Pope F.M., Eady R.A.J.;  
 RT "Clinopathological correlations of compound heterozygous COL7A1  
 RT mutations in recessive dystrophic epidermolysis bullosa."; *J. Invest. Dermatol.* 107:171-177(1996).  
 RL [17]  
 RP VARIANTS RDEB W-1982; G-2008; A-2025; E-2049; G-2063; W-2063 AND  
 RP R-2575.  
 RX MEDLINE-97465605; PubMed-9326325;  
 RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,  
 RA Fraile S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,  
 RA de Prost Y.;  
 RT "Characterization of 18 new mutations in COL7A1 in recessive  
 RT dystrophic epidermolysis bullosa provides evidence for distinct  
 RT molecular mechanisms underlying defective anchoring fibril  
 RT formation."; *Am. J. Hum. Genet.* 61:599-610(1997).  
 RL [18]  
 RP VARIANT RDEB ASP-1652.  
 RX MEDLINE-98106792; PubMed-9444387;  
 RA Cserhalmi-Friedman P.B., Karpaci S., Horvath A., Christiano A.M.;  
 RT "Identification of a glycine substitution and a splice site mutation  
 RT in the type VII collagen gene in a proband with milds recessive  
 RT dystrophic epidermolysis bullosa."; *Arch. Dermatol. Res.* 289:640-645(1997).  
 RL [19]  
 RP VARIANTS DEB ARG-2009 AND ARG-2043.  
 RX MEDLINE-97358588; PubMed-9215684;  
 RA Winberg J.-O., Hammani-Hausaill N., Nilsen O., Anton-Lamprecht I.,  
 RA Naylor S.L., Kerbacher K., Zimmermann M., Krejci P.,  
 RA Gedde-dahl T., Jr., Bruckner-Tuderman L.;  
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by  
 RT a splice site mutation in combination with a missense mutation in the  
 RT COL7A1 gene."; *Hum. Mol. Genet.* 6:1125-1135(1997).  
 RL [20]  
 RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.  
 RX MEDLINE-98334667; PubMed-9668111;  
 RA Hammani-Hausaill N., Schumann H., Ragunath M., Kilgus O., Luechi U.,  
 RA Luger T., Bruckner-Tuderman L.;  
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in  
 RT intracellular accumulation of collagen VII, loss of anchoring

```

RT fibrils, and skin blistering."
RL J. Biol. Chem. 273:19228-19234(1998).
RN [21]
RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
RX MEDLINE=98410969; PubMed=9740253.
RA Kon A., Pulkkinen L., Ishida-Tanamoto A., Hashimoto I., Uitto J.;
RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
RL bullosa."
RN J. Invest. Dermatol. 111:534-537(1998).
RN [22]
RP VARIANT RDEB ARG-1347.
RX MEDLINE=99019477; PubMed=9804332.
RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,
RT Zamburino G., Bruckner-Tuderman L., Castiglia D.;
RT "Compound heterozygosity for a recessive glycine substitution and a
RT splice site mutation in the COL7A1 gene causes an unusually mild form
RL of localized recessive dystrophic epidermolysis bullosa."
RN J. Invest. Dermatol. 111:744-750(1998).
RN [23]
RP VARIANTS DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.
RX MEDLINE=99072663; PubMed=9856843;

Query Match          65.2%; Score 60; DB 1; Length 2944;
Best Local Similarity 66.7%; Pred. No. 1.7;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGPWGPNKGDKGKVG 15
DB 1623 PGPGVGRGRDGEVG 1637
      ||| ||| ||| |||

```

---

```

RESULT 9
CA12_MOUSE          STANDARD;      PRT; 1459 AA.
AC P28481;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(I) chain precursor [Contains: Chondrocalcin].
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91358489; PubMed=1885613;
RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
RL structure, and alternative splicing."
RN J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RP SEQUENCE OF 1455-1459 FROM N.A.
RX MEDLINE=91274355; PubMed=2054384;
RA Metsaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RL collagen mRNAs."
RN Biochim. Biophys. Acta 1089:241-243(1991).
RN -1- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 WMFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB EMBL; M65161; AAA68100.1; -

```

```

DR EMBL; X57982; CAA41047.1; -.
DR MGD; MGI:88452; Col2a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib.collagen_C.
DR InterPro; IPR001007; WMFC.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 17.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLF1; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; WMFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT SIGNAL          1      25
FT PROPEP          26     153
FT CHAIN           154    1213
FT PROPEP          1214    1459
FT DOMAIN           32      89
FT DOMAIN          173    1186
FT DOMAIN          1187    1213
FT VARSPPLIC       29      29
FT VARSPPLIC       30      98
FT VARSPPLIC       30     139154
SQ SEQUENCE 1459 AA; 139154 MW; F6C84FAFC532EF2 CRC64;

Query Match          62.0%; Score 57; DB 1; Length 1459;
Best Local Similarity 73.3%; Pred. No. 2.3;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGPNKGDKGKVG 15
DB 888 PAGPGPGAGKDGPKG 902
      |||| ||| ||| |||

```

---

```

RESULT 10
CA12_CHICK          STANDARD;      PRT; 369 AA.
AC P02460;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(II) chain precursor (Fragment).
GN COL2A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-193 FROM N.A.
RX MEDLINE=85306862; PubMed=3840018;
RA Deak F., Argaves W.S., Kiss I., Sparks K.J., Goetlinc P.F.;
RT "Primary structure of the chicken type II procollagen and a portion of the helical
RT domain of chicken type II procollagen as determined by DNA sequence
RL analysis."
RN Biochem. J. 229:189-196(1985).
RN [2]
RP SEQUENCE OF 82-369 FROM N.A.
RX MEDLINE=84239728; PubMed=6330084;
RA Sandell L.J., Prentice H.L., Kravits D., Upholt W.B.;
RT "Structure and sequence of the chicken type II procollagen gene.
RT Characterization of the region encoding the carboxyl-terminal
RT telopeptide and propeptide."
RN J. Biol. Chem. 259:7826-7834(1984).
RN [3]
RP SEQUENCE OF 114-369 FROM N.A.
RX Niimiya Y., Shewalter A.M., van der Rest M., Seldah N.G.,
RA Chretien M., Olsen B.R.;
RT "Structure of the carboxyl propeptide of chicken type II procollagen
RT determined by DNA and protein sequence analysis."
RL Biochemistry 23:617-624(1984).

```

```

CC -1- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02663; CAA26499.1; -.
DR EMBL: L00063; AAB59967.1; JOINED.
DR EMBL: L00061; AAB59967.1; JOINED.
DR EMBL: L00062; AAB59967.1; JOINED.
DR PIR: A02860; CCGH6C.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib.collagen_C.
DR InterPro: IPR001007; VMFC.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 1.
DR ProDom: PD002078; Fib.collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR PROSITE: PS01208; VMFC; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Cartilage; Collagen.
FT NON_TER 1
FT CHAIN <1 123 COLLAGEN ALPHA 1(II) CHAIN.
FT PROPEP 124 369 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN <1 96 TRIPLE-HELICAL REGION.
FT DOMAIN 97 123 NONHELICAL REGION (C-TERMINAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .).
FT DISULFID 275 320
SQ SEQUENCE 369 AA; 38989 MW; EF5306925B0BA380 CRC64;

Query Match 60.9%; Score 56; DB 1; Length 369;
Best Local Similarity 66.7%; Pred. No. 0.88;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPWGPGKDGKVG 15
Db 47 PPGVGPSPGKDGSGNG 61

RESULT 11
CA13_RAT STANDARD; PRT; 636 AA.
ID CA13_RAT
AC P13941; 070604;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain (Fragment).
GN COL3A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94114571; Pubmed=8286415;
RA Glumoff V., Macelae J.K., Vuorio E.;
RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different
RT expression patterns of type I and type III collagen and fibronectin
RT genes in experimental granulation tissue."
RL Biochim. Biophys. Acta 1217:41-48(1994).
RN [2]
RP SEQUENCE OF 73-636 FROM N.A.
RC STRAIN-Sprague-Dawley; TISSUE-Fibroblast;
RA Wurtz T., Ellerstrom C., Lundmark C., Christersson C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]

```

```

RP SEQUENCE OF 308-482 FROM N.A.
RX MEDLINE=88296083; Pubmed=2456904;
RA Frankel F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lytle C.R.,
RA Komm B., Mohr K.;
RT "Regulation of alpha 2(I), alpha 1(III), and alpha 2(V) collagen
RT mRNAs by estradiol in the immature rat uterus."
RL DNA 7:347-354(1988).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSSED-LINKED VIA HYDROXYLYSINS.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X70369; CAA49832.1; -.
DR EMBL: AJ005395; CAA06510.1; -.
DR EMBL: M21354; AAA40942.1; -.
DR PIR: A29905; A29905.
DR PIR: S41067; S41067.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib.collagen_C.
DR InterPro: IPR001007; VMFC.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 6.
DR ProDom: PD002078; Fib.collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR PROSITE: PS01208; VMFC; PARTIAL.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Glycoprotein.
FT NON_TER 1
FT CHAIN <1 375 COLLAGEN ALPHA 1(III) CHAIN.
FT PROPEP 376 636 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN <1 368 TRIPLE-HELICAL REGION.
FT DOMAIN 369 636 NONHELICAL REGION (C-TERMINAL).
FT DISULFID 368 368 INTERCHAIN (BY SIMILARITY).
FT DISULFID 369 369 INTERCHAIN (BY SIMILARITY).
FT CONFLICT 340 340 N->D (IN REF. 2).
FT CONFLICT 429 429 A->G (IN REF. 2).
SQ SEQUENCE 636 AA; 62332 MW; 61A48159F01D01EE CRC64;

Query Match 60.9%; Score 56; DB 1; Length 636;
Best Local Similarity 66.7%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWGPGKDGKVG 15
Db 313 PVPGPSPGKDGSSG 327

RESULT 12
CA1B_MOUSE STANDARD; PRT; 1804 AA.
ID CA1B_MOUSE
AC 061245; 064047;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(XI) chain precursor.
GN COL11A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

CC TISSUE-Embryo; PubMed=8530046;  
 CC MEDLINE=96015067; PubMed=8530046;  
 RA Yoshio H., Inoguchi K., Khaleduzaman M., Nimomiya Y.,  
 RA Andrikopoulos K., Ramirez F.;  
 RT "Coding sequence and alternative splicing of the mouse alpha 1(XI)  
 RT collagen gene (Col1a1)";  
 RL Genomics 28:337-340(1995).  
 RN [2]  
 RP SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=95163095; PubMed=7859283;  
 RA Li Y., Lacerda D.A., Warman M.L., Beler D.R., Yoshio H.,  
 RA Nimomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,  
 RA Ramirez F., Wardell B.B., Liferth G.D., Teuscher C., Woodward S.R.,  
 RA Taylor B.A., Seegmiller R.E., Olsen B.R.;  
 RT "A fibrillar collagen gene, Col1a1, is essential for skeletal  
 RT morphogenesis";  
 RL Cell 80:423-430(1995).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)-1(II) (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM, ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DISEASE: DEFECTS IN COL1A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA,  
 CC AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS  
 CC CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES  
 CC AND TRACHEA.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D38162; BAA07367.1; -;  
 DR EMBL: S74574; AAB33439.1; -;  
 DR MGD: MGI:88446; Coll1a1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001791; Lamlna1n\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Pfam: PF02210; TSPN; 1.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI; 1.  
 DR SMART: SM00282; Lamg; 1.  
 DR SMART: SM00210; TSPN; 1.  
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Collagen; Signal; Alternative splicing;  
 KM Disease mutation.  
 FT SIGNAL 1 35  
 FT PROPEP 36 511 POTENTIAL.  
 FT CHAIN 512 1561 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).  
 FT PROPEP 1562 1804 COLLAGEN ALPHA 1(XI) CHAIN.  
 FT DOMAIN 37 417 CARBOXY-TERMINAL PROPEPTIDE.  
 FT DOMAIN 418 506 NONHELICAL REGION.  
 FT DOMAIN 507 509 TRIPLE-HELICAL REGION (INTERRUPTED).  
 FT DOMAIN 510 527 SHORT NONHELICAL SEGMENT.  
 FT DOMAIN 528 1540 TELOPEPTIDE.  
 FT DOMAIN 1541 1561 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1561 1561 TRIPLE-HELICAL REGION (C-TERMINAL).  
 FT CARBOHD 1638 1638 NONHELICAL REGION (C-TERMINAL).  
 FT SITE 610 610 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 1450 1450 CROSSLINKING.  
 FT SITE 1450 1450 CROSSLINKING.

FT VARSLIC 329 413 MISSING (IN SHORT ISOFORM).  
 FT VARIANT 189 195 IDRSERS -> SIEVDR (IN CHONDRODYSPLASIA).  
 FT VARIANT 196 1804 MISSING (IN CHONDRODYSPLASIA).  
 SQ SEQUENCE 1804 AA; 180963 MW; FE2DB9DED14219A CRC64;  
 Query Match 60.9%; Score 56; DB 1; Length 1804;  
 Best Local Similarity 66.7%; Pred. No. 3.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PAGMGPGNGDKVKG 15  
 Db 1224 PRPGGPGNADGPGQ 1238  
 ID CALB\_HUMAN STANDARD; PRT; 1806 AA.  
 AC P12107; Q14034;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(XI) chain precursor.  
 GN COL1A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90202924; PubMed=1690726;  
 RA Yoshio H., Ramirez F.;  
 RT "Pro-alpha 1(XI) collagen. Structure of the amino-terminal propeptide  
 RT and expression of the gene in tumor cell lines";  
 RL J. Biol. Chem. 265:6423-6426(1990).  
 RN [2]  
 RP SEQUENCE OF 538-1806 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=89034222; PubMed=3182841;  
 RA Bernard M., Yoshio H., Rodriguez E., van der Rest M., Kimura T.,  
 RA Nimomiya Y., Olsen B., Ramirez F.;  
 RT "Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA  
 RT demonstrates that type XI belongs to the fibrillar class of collagens  
 RT and reveals that the expression of the gene is not restricted to  
 RT cartilaginous tissue";  
 RL J. Biol. Chem. 263:17159-17166(1988).  
 RN [3]  
 RP ALTERNATIVE SPLICING.  
 RC TISSUE=Blood;  
 RX MEDLINE=95238468; PubMed=7721876;  
 RA Zhidkova N.I., Justice S.K., Wayne R.;  
 RT "Alternative mRNA processing occurs in the variable region of the  
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains";  
 RL J. Biol. Chem. 270:9486-9493(1995).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)-1(II).  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING. THERE IS ALTERNATIVE USAGE OF  
 CC EXON IIA OR EXON IIB. BOTH EXONS IIA AND IIB ARE PRESENT IN  
 CC TRANSCRIPTS FROM CARTILAGE BUT EXON IIB IS PREFERENTIALLY UTILIZED  
 CC IN TRANSFORMED CELL LINES.  
 CC -1- TISSUE SPECIFICITY: CARTILAGE, PLACENTA AND SOME TUMOR OR VIRALLY  
 CC TRANSFORMED CELL LINES.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J04177; AA51891.1; -  
 DR EMBL: L38956; AAA79171.1; -  
 DR PIR: A31795; A31795.  
 DR PIR: A35239; A35239.  
 DR MIM: 120280; -  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01410; COLFI.1.  
 DR Pfam: PF01391; Collagen.18.  
 DR Pfam: PF02210; TSPN.1.  
 DR ProDom: PD002078; Fib\_collagen\_C.1.  
 DR SMART: SM00038; COLFI.1.  
 DR SMART: SM00282; LamG.1.  
 DR SMART: SM00210; TSPN.1.  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal; Alternative splicing.  
 FT SIGNAL 1 36  
 FT PROPEP 37 511  
 FT CHAIN 512 1563  
 FT PROPEP 1564 1806  
 FT DOMAIN 37 419  
 FT DOMAIN 420 508  
 FT DOMAIN 509 511  
 FT DOMAIN 512 528  
 FT DOMAIN 529 1542  
 FT DOMAIN 1543 1563  
 FT CARBOHYD 1640 1640  
 FT SITE 612 612  
 FT SITE 1452 1452  
 FT VARSPLIC 261 299

SO SEQUENCE 1806 AA; 181137 MW; C27AD86180BCDEF5 CRC64;  
 Query Match 60.9%; Score 56; DB 1; Length 1806;  
 Best Local Similarity 66.7%; Score No. 3.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAGEWPGNGDKAYG 15  
 Db 1226 PRGPGGNGADGPG 1240

RESULT 14  
 ID CA13\_CHICK STANDARD; PRT; 1262 AA.  
 AC P12105; P79758; P79759; Q90794; Q92029;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(III) chain precursor (fragments).  
 GN COL3A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCB1;TaxID=9031;  
 NC NCB1;TaxID=9031;  
 RN [1]  
 RP SEQUENCE OF 1-886 FROM N.A.  
 RC TISSUE-Kidney;  
 RX MEDLINE=94265842; PubMed=8206952;  
 RA Nish H.-D., Niu Z., Adams S.L.;  
 RT "An alternative transcript of the chick type III collagen gene that  
 RL does not encode type III collagen.";  
 RL J. Biol. Chem. 269:16443-16448(1994).

RN [2]  
 RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.  
 RX MEDLINE=84270696; PubMed=6547770;  
 RA Yamada Y., Liao G., Mudryj M., Obici S., de Crombrughe B.;  
 RT "Conservation of the sizes for one but not another class of exons in  
 RL two chick collagen genes.";  
 RL Nature 310:333-337(1984).  
 RN [3]  
 RP SEQUENCE OF 977-1262 FROM N.A.  
 RX MEDLINE=83220816; PubMed=6856474;  
 RA Yamada Y., Kuhn K., de Crombrughe B.;  
 RT "A conserved nucleotide sequence, coding for a segment of the C-  
 RL propeptide, is found at the same location in different collagen  
 RL genes.";  
 CC Nucleic Acids Res. 11:2733-2744(1983).  
 CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC ALONG WITH TYPE I COLLAGEN.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLISINES.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: CONTAINS 1 VMFC DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U07973; AAA83407.1; -  
 DR EMBL: X00822; CAB52686.1; -  
 DR EMBL: X00823; CAB52686.1; JOINED.  
 DR EMBL: X00826; CAA25397.1; ALT\_SEQ.  
 DR EMBL: X00825; CAA25397.1; JOINED.  
 DR EMBL: X00827; CAA25398.1; -  
 DR EMBL: X00828; CAA25399.1; -  
 DR EMBL: X00830; CAA25401.1; -  
 DR EMBL: X00831; CAA25402.1; -  
 DR EMBL: X00831; AAD15299.1; -  
 DR EMBL: K02302; AAD15299.1; -  
 DR EMBL: K02301; AAD15298.1; -  
 DR EMBL: M36662; AAA18519.1; ALT\_SEQ.  
 DR PIR: A05269; A05269.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR InterPro: IPR001007; VMFC.  
 DR Pfam: PF01391; Collagen.13.  
 DR ProDom: PD002078; Fib\_collagen\_C.1.  
 DR SMART: SM00038; COLFI.1.  
 DR SMART: SM00214; VMC.1.  
 DR PROSITE: PS01208; VMC.1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 23  
 FT PROPEP 24 144  
 FT CHAIN 145 1003  
 FT PROPEP 1004 1262  
 FT DOMAIN 29 88  
 FT DOMAIN 145 164  
 FT DOMAIN 165 994  
 FT DOMAIN 995 1003  
 FT NON CONS 886 887  
 FT NON CONS 922 923  
 FT DISULFID 994 994  
 FT DISULFID 995 995  
 FT MOD\_RES 262 262  
 FT MOD\_RES 283 283

INTERCHAIN (BY SIMILARITY).  
 INTERCHAIN (BY SIMILARITY).  
 HYDROXYLATION (BY SIMILARITY).  
 HYDROXYLATION (BY SIMILARITY).  
 HYDROXYLATION (BY SIMILARITY).  
 HYDROXYLATION (BY SIMILARITY).  
 AMINO-TERMINAL PROPEPTIDE (BY  
 SIMILARITY).  
 COLLAGEN ALPHA 1(III) CHAIN.  
 CARBOXYL-TERMINAL PROPEPTIDE (BY  
 SIMILARITY).  
 VMFC.  
 NONHELICAL REGION (N-TERMINAL) (BY  
 SIMILARITY).  
 TRIPLE-HELICAL REGION (BY SIMILARITY).  
 NONHELICAL REGION (C-TERMINAL) (BY  
 SIMILARITY).

FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
FT CARBOHYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 96 96 E -> K (IN REF. 2).  
FT CONFLICT 1132 1132 F -> S (IN REF. 3).  
SQ SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;  
  
Query Match 59.8%; Score 55; DB 1; Length 1262;  
Best Local Similarity 66.7%; Pred. No. 3.7;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 PAGEWPGNCKGCKVG 15  
DB 942 PPGPAGPEKDGKRG 956  
  
RESULT 15  
CA21\_CHICK STANDARD; PRT: 1362 AA;  
ID CA21\_CHICK STANDARD; PRT: 1362 AA;  
AC P02467; Q90795; Q90797; P87492; Q90758; Q92014; P87491; Q90792;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Collagen alpha 2(I) chain precursor (Fragments).  
OS COL1A2.  
GN Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.  
RX MEDLINE=86185168; PubMed=3868961;  
RA Boedtker H., Finer M., Aho S.;  
RL "The structure of the chicken alpha 2 collagen gene.";  
RT Ann. N.Y. Acad. Sci. 460:85-116(1985).  
RN [2]  
RP SEQUENCE OF 1-89 FROM N.A.  
RX MEDLINE=83246518; PubMed=6135195;  
RA Tate V.E., Finer M.H., Boedtker H., Doty P.;  
RL "Chick pro alpha 2 (I) collagen gene: exon location and coding  
potential for the prepropeptide.";  
RN [3]  
RL Nucleic Acids Res. 11:91-104(1983).  
RN [4]  
RP SEQUENCE OF 1-14 FROM N.A.  
RX MEDLINE=82060240; PubMed=6946474;  
RA Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,  
RL de Crombrughe B.;  
RN [5]  
RP "Structure of the promoter for chicken alpha 2 type I collagen gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).  
RN [6]  
RP SEQUENCE OF 1-33 FROM N.A.  
RX MEDLINE=84297217; PubMed=6473103;  
RA Aho S., Tate V.E., Boedtker H.;  
RL "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen  
gene.";  
RN [7]  
RL Nucleic Acids Res. 12:6117-6125(1984).  
RN [8]  
RP SEQUENCE OF 1-79 FROM N.A.  
RX MEDLINE=88056316; PubMed=3678834;  
RA Finer M.H., Boedtker H., Doty P.;  
RL "Construction and characterization of cDNA clones encoding the 5' end  
of the chicken pro alpha 1(I) collagen mRNA.";  
RN [9]  
RP Gene 56:71-78(1987).  
RN [10]  
RP SEQUENCE OF 78-92.  
RC TISSUE=Skin;  
RX MEDLINE=71115216; PubMed=5544653;  
RA Higberger J.H., Kang A.H., Gross J.;  
RL "Comparative studies on the amino acid sequence of the alpha 2-CB2  
peptides from chick and rat skin collagens.";  
RN [11]  
RL Biochemistry 10:610-616(1971).  
RN [12]  
RP SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.

RX MEDLINE=82058081; PubMed=6272119;  
RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;  
RL "Structure of the pro alpha 2 (I) collagen gene.";  
RN [13]  
RP Nature 294:129-135(1981).  
RN [14]  
RP SEQUENCE OF 78-92.  
RC TISSUE=Skin;  
RX MEDLINE=70131186; PubMed=4313735;  
RA Kang A.H., Gross J.;  
RL "Amino acid sequence of cyanogen bromide peptides from the amino-  
terminal region of chick skin collagen.";  
RN [15]  
RL Biochemistry 9:796-804(1970).  
RN [16]  
RP SEQUENCE OF 78-92 AND 415-448.  
RC TISSUE=Skin;  
RX MEDLINE=69285369; PubMed=5809220;  
RA Kang A.H., Igarashi S., Gross J.;  
RL "Characterization of the cyanogen bromide peptides from the alpha-2  
chain of chick skin collagen.";  
RN [17]  
RL Biochemistry 8:3200-3204(1969).  
RN [18]  
RP SEQUENCE OF 78-92 AND 415-448.  
RC TISSUE=Bone;  
RX MEDLINE=69206882; PubMed=5785233;  
RA Lane J.M., Miller E.J.;  
RL "Isolation and characterization of the peptides derived from the  
alpha 2 chain of chick bone collagen after cyanogen bromide  
cleavage.";  
RN [19]  
RL Biochemistry 8:2134-2139(1969).  
RN [20]  
RP SEQUENCE OF 566-587 FROM N.A.  
RX MEDLINE=79074829; PubMed=364479;  
RA Lehnach H., Frieschaut A.-M., Hanahan D., Wozney J., Fuller F.,  
RL Kravenchak R., Boedtker H., Doty P.;  
RN "Construction and characterization of a 2.5-kilobase procollagen  
clone.";  
RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).  
RN [21]  
RP SEQUENCE OF 902-1362 FROM N.A.  
RX MEDLINE=81160715; PubMed=6927845;  
RA Fuller F., Boedtker H.;  
RL "Sequence determination and analysis of the 3' region of chicken pro-  
alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
including the carboxy-terminal propeptide sequences.";  
RN [22]  
RL Biochemistry 20:996-1006(1981).  
RN [23]  
RP SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.  
RX MEDLINE=81264246; PubMed=6267043;  
RA Dickson L.A., Nihomiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,  
RL Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,  
RA Fietzek P.P., Olsen B.R.;  
RL "The exon/intron structure of the 3'-region of the pro alpha 2(I)  
collagen gene.";  
RN [24]  
RL J. Biol. Chem. 256:8407-8415(1981).  
RN [25]  
RP SEQUENCE OF 932-954 AND 968-980 FROM N.A.  
RX MEDLINE=81064671; PubMed=6159982;  
RA Avvedimento V.E., Vogeli G., Yamada Y., Matzel J.V. Jr., Pastan I.,  
RL de Crombrughe B.;  
RN "Correlation between splicing sites within an intron and their  
sequence complementarity with UI RNA.";  
RN [26]  
RL Cell 21:689-696(1980).  
RN [27]  
RP SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.  
RX MEDLINE=81112157; PubMed=7460017;  
RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,  
RL Imani M., Pastan I., de Crombrughe B.;  
RN "The collagen gene: evidence for its evolutionary assembly by  
RT amplification of a DNA segment containing an exon of 54 bp.";  
RL Cell 22:887-892(1980).  
RN [28]  
RP -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
CC (FIBRILLAR FORMING COLLAGEN).  
CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.





RT the human type II procollagen gene.";  
RL Nucleic Acids Res. 12:1025-1038(1984).  
[9]  
RP SEQUENCE OF 35-167 FROM N.A.  
RX MEDLINE=89233138; PubMed=2714801;  
RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;  
RT "Organization of the exons coding for pro alpha 1(I) collagen N-  
RT propeptide confirms a distinct evolutionary history of this domain of  
RL the fibrillar collagen genes.";  
RL Genomics 4:438-441(1989).  
[10]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=91184577; PubMed=2010058;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in collagen genes: causes of rare and some common diseases  
RT in humans.";  
RL FASEB J. 5:2052-2060(1991).  
[11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
RT associated collagen (type IX), and network-forming collagen (type X)  
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
[12]  
RP VARIANT SER-1074.  
RX MEDLINE=90036909; PubMed=2572591;  
RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,  
RA Hollister D.W.;  
RT "Glycine to serine substitution in the triple helical domain of pro-  
RT alpha 1 (II) collagen results in a lethal perinatal form of short-  
RT limbed dwarfism.";  
RL J. Biol. Chem. 264:18265-18267(1989).  
[13]  
RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.  
RX MEDLINE=89266907; PubMed=2543071;  
RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;  
RT "Identification of the molecular defect in a family with  
RT spondyloepiphyseal dysplasia.";  
RL Science 244:978-980(1989).  
[14]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE=90370826; PubMed=1975693;  
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;  
RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
RT cause of primary osteoarthritis associated with a mild  
RT chondrodysplasia.";  
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).  
[15]  
RP VARIANT OI-IV VAL-717.  
RX MEDLINE=91291136; PubMed=2064612;  
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;  
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
RT valine substitution in osteogenesis imperfecta type IV. Detection of  
RT the mutation and prenatal diagnosis by a chemical cleavage method.";  
RL Biochem. J. 276:765-770(1991).  
[16]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE=91086471; PubMed=1985108;  
RA Eyre D.R., Weis M.A., Moskowitz R.W.;  
RT "Cartilage expression of a type II collagen mutation in an inherited  
RT form of osteoarthritis associated with a mild chondrodysplasia.";  
RL J. Clin. Invest. 87:357-361(1991).  
[17]  
RP VARIANT HYPOCHONDROGENESIS GLU-984.  
RX MEDLINE=93054548; PubMed=1429602;  
RA Bogaert R., Tiller G.E., Wiles M.A., Gruber H.E., Rimoin D.L.,  
RA Cohn D.H., Eyre D.R.;  
RT "An amino acid substitution (Gly983-->Glu) in the collagen alpha  
RT 1(I) chain produces hypochondrogenesis.";  
RL J. Biol. Chem. 267:22522-22526(1992).  
[18]  
RP VARIANT HYPOCHONDROGENESIS SER-705.  
RX MEDLINE=92262484; PubMed=1374906;  
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,  
RA Ramlier F., Vitale E., Lee B.;  
RT "Characterization of a type II collagen gene (COL2A1) mutation  
RT identified in cultured chondrocytes from human hypochondrogenesis.";  
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).  
[19]  
RP VARIANT WS-II ASP-198.  
RX MEDLINE=93304428; PubMed=8317498;  
RA Koerike J., Ritvaniemi P., Haataja L., Kaeerlaeinen H.,  
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;  
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
RT for glycine alpha 1-67 and that causes cataracts and retinal  
RT detachment: evidence for molecular heterogeneity in the Wagner  
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";  
RL Am. J. Hum. Genet. 53:55-61(1993).  
[20]  
RP VARIANT SEDC CYS-840.  
RA Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,  
RA Eyre D.R.;  
RT "A dominant mutation in the type II collagen gene (COL2A1) produces  
RT spondyloepimetaphyseal dysplasia (SEMD), Strudwick type.";  
RL Am. J. Hum. Genet. 53:A209-A209(1993).  
[21]  
RP VARIANT OSTEOARTHRITIS CYS-650.  
RX MEDLINE=93282819; PubMed=8507190;  
RA Holdreham D., Malmud C.J., Moskowitz R.W., Haqqi T.M.;  
RT "Human cartilage from late stage familial osteoarthritis transcribes  
RT type II collagen mRNA encoding a cysteine in position 519.";  
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).  
[22]  
RP VARIANT SEDC ARG-285.  
RX MEDLINE=93252400; PubMed=8486375;  
RA Viikula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,  
RA Peltonen L.;  
RT "A mutation in the amino-terminal end of the triple helix of type II  
RT collagen causing severe osteochondrodysplasia.";  
RL Genomics 16:282-285(1993).  
[23]  
RP VARIANT SEDC CYS-206.  
RX MEDLINE=94063862; PubMed=8244341;  
RA Williams C.J., Constidine E.L., Knowlton R.G., Reginato A., Neumann G.,  
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;  
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a  
RT family with an Arg75-->Cys mutation in the procollagen type II gene  
RL (COL2A1).";  
RL Hum. Genet. 92:499-505(1993).  
[24]  
RP VARIANT SEDC CYS-920.  
RX MEDLINE=93315508; PubMed=8325895;  
RA Chan D., Taylor T.K.F., Cole W.G.;  
RT "Characterization of an arginine 789 to cysteine substitution in  
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal  
RT dysplasia.";  
RL J. Biol. Chem. 268:15238-15245(1993).  
[25]  
RP VARIANT SEDC SER-1128.  
RX MEDLINE=93140139; PubMed=8423604;  
RA Cole W.G., Hall R.K., Rogers J.G.;  
RT "The clinical features of spondyloepiphyseal dysplasia congenita  
RT resulting from the substitution of glycine 997 by serine in the alpha  
RT 1(II) chain of type II collagen.";  
RL J. Med. Genet. 30:27-35(1993).  
Query Match 59.8%; Score 55; DB 1; Length 1418;  
Best Local Similarity 66.7%; Pred. NO. 4.2;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
OY 1 PAGPBGNGKXGVG 15.  
DB 1096 PGPVPGSGKDGANG 1110





```

RN [2]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogel G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RT alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [3]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liu G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [4]
RP SEQUENCE OF 810-1464 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic head;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [5]
RP SEQUENCE OF 1442-1464 FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=91274355; PubMed=2054384;
RA Metsaenta M., Toman D., de Crombrughe B., Viorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs.";
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCAN CONSISTS OF GIC-GAL DISACCHARIDE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 WPC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52046; CAA36279.1; -
DR EMBL: M18833; AAA37338.1; -
DR EMBL: K03037; -; NOT_ANNOTATED_CDS.
DR EMBL: AK019448; BAB31724.1; -
DR EMBL: X57983; CAA41048.1; -
DR PIR: A22287; A22287.
DR PIR: A27353; A27353.
DR PIR: S16373; S16373.
DR MGI: 88453; Col3a1.

```

```

DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; WPC.
DR Pfam; PF01410; COLFI. 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WPC; 1.
DR ProSite; PS01208; WPC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 23
FT PROPEP 24 154
FT CHAIN 155 1203
FT PROPEP 1204 1464
FT DOMAIN 31 90
FT DOMAIN 155 169
FT DOMAIN 170 1195
FT DOMAIN 1196 1464
FT CARBOHYD 262 262
FT MOD_RES 262 262
FT MOD_RES 283 283
FT MOD_RES 859 859
FT MOD_RES 976 976
FT MOD_RES 1093 1093
FT MOD_RES 1105 1105
FT DISULFD 1196 1196
FT SEQUENCE 1464 AA; 138944 MW; 210AEC27A886090B CRC64;
SQ
Query Match 59.8%; Score 55; DB 1; Length 1464;
Best Local Similarity 66.7%; Pred. No. 4.3;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 PAGPWGPNKGDKVG 15
Db 1140 PVGPHGPEKDGTS 1154
ID CAA2_MOUSE STANDARD; PRT: 1650 AA.
AC Q6439; Q921W0;
DT 15-JUL-1998 (Rel. 36, Created)
DR 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL1A2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region.";
RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-1592 FROM N.A.
RC STRAIN=FVB/N, AND 129SV; TISSUE=Cartilage;
RX MEDLINE=97135795; PubMed=8981332;
RA Vandenberg P., Vioristo M.M., Ala-Kokko L., Prockop D.J.;
RT "The mouse colla2 gene. Some transcripts from the adjacent rrx-beta
RT gene extend into the colla2 gene.";
RL Matrix Biol. 15:359-367(1996).
CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY
CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),
CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL
CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD

```

CC OF ALPHA 3(XI)-1(II) (BY SIMILARITY).  
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS ARE PRODUCED BY  
 CC -1- ALTERNATIVE SPLICING.  
 CC -1- PPM: PROLINS AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF100956; AAC69905.1; -  
 DR EMBL: U16789; AAA67751.1; -  
 DR EMBL: U16790; AAA67752.1; -  
 DR MCD: MGI:88447; COL11A2.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fibcollagen\_C.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01410; COLF1; 2.  
 DR Pfam: PF02210; TSPN; 1.  
 DR ProDom: PD002078; Fibcollagen\_C; 1.  
 DR SMART: SM00038; COLF1; 1.  
 DR SMART: SM00282; Lamc; 1.  
 DR SMART: SM00210; TSPN; 1.  
 KM Extracellular matrix; Collagen; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Alternative splicing; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 1414  
 FT PROPEP 1415 1650  
 FT DOMAIN 23 255  
 FT DOMAIN 487 1414  
 FT DOMAIN 1415 1650  
 FT CONFLICT 450 450  
 FT CONFLICT 618 619  
 FT CONFLICT 711 711  
 FT CONFLICT 757 759  
 FT CONFLICT 768 768  
 FT CONFLICT 790 790  
 FT CONFLICT 803 803  
 FT CONFLICT 836 836  
 FT CONFLICT 919 919  
 FT CONFLICT 1167 1167  
 FT CONFLICT 1300 1300  
 FT CONFLICT 1436 1436  
 SQ SEQUENCE 1650 AA; 162072 MW; 8ECD8B702E71E2DA CRC64;  
 Query Match 59.8%; Score 55; DB 1; Length 1650;  
 Best Local Similarity 66.7%; Pred. No. 4.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-96032717; PubMed-7559422;  
 RA Vioristo M.M., Pihlajama T., Vandenberg P., Prockop D.J.,  
 RA Ala-Kokko L.;  
 RT "The human COL1A2 gene structure indicates that the gene has not  
 RT evolved with the genes for the major fibrillar collagens.";  
 RL J. Biol. Chem. 270:22873-22881(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Tubby B.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 59-807 FROM N.A.  
 RC TISSUE-Cartilage;  
 RX MEDLINE-93314796; PubMed-8325374;  
 RA Zhidkova N.I., Brewton R.G., Mayne R.;  
 RT "Molecular cloning of PARP (proline/arginine-rich protein) from human  
 RT cartilage and subsequent demonstration that PARP is a fragment of the  
 RT NH2-terminal domain of the collagen alpha 2(XI) chain.";  
 RL FEBS Lett. 326:25-28(1993).  
 RN [4]  
 RP SEQUENCE OF 730-1690 FROM N.A.  
 RX MEDLINE-89340485; PubMed-2760050;  
 RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,  
 RA van der Rest M., Ono K., Solomon E., Nishimura Y., Olsen B.R.;  
 RT "The human alpha 2(XI) collagen (COL1A2) chain. Molecular cloning of  
 RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen  
 RT with differences in genomic organization.";  
 RL J. Biol. Chem. 264:13910-13916(1989).  
 RN [5]  
 RP SEQUENCE OF 1-537 FROM N.A.  
 RX MEDLINE-96435918; PubMed-8838804;  
 RA Lui V.C., Ng L.J., Sat E.W., Cheah K.S.;  
 RT "The human alpha 2(XI) collagen gene (COL1A2): completion of coding  
 RT information, identification of the promoter sequence, and precise  
 RT localization within the major histocompatibility complex reveal  
 RT overlap with the KES gene.";  
 RL Genomics 32:401-412(1996).  
 RN [6]  
 RP ALTERNATIVE SPLICING.  
 RX MEDLINE-95238468; PubMed-7721876;  
 RA Zhidkova N.I., Justice S.K., Mayne R.;  
 RT "Alternative mRNA processing occurs in the variable region of the  
 RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";  
 RL J. Biol. Chem. 270:9486-9493(1995).  
 RN [7]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE-97255959; PubMed-9101290;  
 RA Kulvanenti H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 RT associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [8]  
 RP VARIANT ARG-661.  
 RX MEDLINE-95163096; PubMed-7859284;  
 RA Viikula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,  
 RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,  
 RA van den Hoogen F.H.J., Ropers C.W.R.J., Takano M., Li S.-W.,  
 RA Olsen B.R., Warman M.L., Brunner H.G.;  
 RT "Autosomal dominant and recessive osteochondrodysplasias associated  
 RT with the COL1A2 locus.";  
 RL Cell 80:431-437(1995).  
 RN [9]  
 RP VARIANTS DFN13 GLU-808 AND CYS-1034, AND REVISIONS TO 1031-1032.  
 RX MEDLINE-20047768; PubMed-10581026;  
 RA McGuire W.T., Prasad S.D., Griffith A.J., Kunst H.P.M., Green G.E.,  
 RA Shpargel K.B., Runge G., Huybrechts C., Mueller R.F., Lynch E.,  
 RA King M.-C., Brunner H.G., Cremers C.W.R.J., Takano M., Li S.-W.,  
 RA Arita M., Mayne R., Prockop D.J., Van Camp G., Smith R.J.H.;  
 RT "Mutations in COL1A2 cause non-syndromic hearing loss (DFN13).";

RL Nat. Genet. 23:413-419(1999).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD  
 CC OF ALPHA 3(XI)-1(II).  
 CC -1- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING. THESE FORMS DIFFER IN THE  
 CC PRESENCE OR ABSENCE OF THREE SEPARATE REGIONS IN THE N-TERMINAL  
 CC NON-COLLAGENOUS DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH  
 CC PEPTIDE OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING  
 CC EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE  
 CC CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT  
 CC AMOUNTS.  
 CC -1- DISEASE: DEFECTS IN COL11A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT  
 CC AND RECESSIVE OSTEOCHONDRODYSPLASIAS. (1) THE AUTOSOMAL DOMINANT  
 CC FORM OF STICKLER SYNDROME (SS) IS CHARACTERIZED BY MILD  
 CC SPONDYLOEPHRYSEAL DYSPLASIA, OSTEOARTHRITIS, AND SENSORINEURAL  
 CC HEARING LOSS. (2) THE AUTOSOMAL RECESSIVE DISORDER IS SIMILAR BUT  
 CC MORE SEVERE.  
 CC -1- DISEASE: DEFECTS IN COL11A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT  
 CC NONSYNDROMIC SENSORINEURAL DEAFNESS TYPE 13 (DFNA13). AFFECTED  
 CC INDIVIDUALS EXPERIENCE PROGRESSIVE HEARING LOSS BEGINNING IN THE  
 CC SECOND TO FOURTH DECADES, EVENTUALLY MAKING USE OF AMPLIFICATION  
 CC MANDATORY.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.  
 CC -1- DATABASE: NAME-Hereditary hearing loss homepage:  
 CC NOTE-Gene page:  
 CC WWW="http://www.uia.ac.be/dnalab/hhh/hhgenes.html".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).  
 CC -----  
 CC  
 DR EMBL: U32169; AAC50214.1; -;  
 DR EMBL: U32169; AAC50213.1; -;  
 DR EMBL: U32169; AAC50215.1; -;  
 DR EMBL: AL031228; CAA20240.1; -;  
 DR EMBL: L18987; AAA35498.1; -;  
 DR EMBL: J04974; AAA52034.1; -;  
 DR EMBL: U41069; AAC17464.1; -;  
 DR EMBL: U41065; AAC17464.1; JOINED.  
 DR EMBL: U41065; AAC17464.1; JOINED.  
 DR EMBL: U41067; AAC17464.1; JOINED.  
 DR PIR: A32645; A32645.  
 DR MIM: 130290; -;  
 DR MIM: 108300; -;  
 DR MIM: 215150; -;  
 DR MIM: 601868; -;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib.collagen\_C.  
 DR InterPro: IPR001791; Lamlnln\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01410; COLF1\_2.  
 DR Pfam: PF01391; Collagen\_18.  
 DR Pfam: PF02210; TSPN\_1.  
 DR ProDom: PD002078; Fib.collagen\_C\_1.  
 DR SMART: SM00038; COLF1\_1.  
 DR SMART: SM00282; LamG\_1.  
 DR SMART: SM00210; TSPN\_1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Alternative splicing; Signal;  
 KW Disease mutation; Deafness.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT CHAIN 23 1500 COLLAGEN ALPHA 2(XI) CHAIN.

FT PROPEP 1501 1736 CARBOXYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 23 255 NONHELICAL REGION.  
 FT DOMAIN 487 1500 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1501 1736 NONHELICAL REGION.  
 FT DOMAIN 10 18 POLY-LEU.  
 FT DOMAIN 298 301 POLY-GLU.  
 FT CARBOHYD 1604 1604 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 267 292 MISSING (IN ISOFORMS WITHOUT EXON 6).  
 FT VARSPLIC 293 313 MISSING (IN ISOFORMS WITHOUT EXON 7).  
 FT VARSPLIC 314 373 MISSING (IN ISOFORMS WITHOUT EXON 8).  
 FT VARIANT 661 661 G -> R (IN AUTOSOMAL RECESSIVE  
 FT OSTEOCHONDRODYSPLASIA).  
 FT /FTId-VAR\_001907.  
 FT VARIANT 808 808 G -> E (IN DFNA13).  
 FT /FTId-VAR\_010655.  
 FT VARIANT 1034 1034 R -> C (IN DFNA13).  
 FT /FTId-VAR\_010656.  
 FT CONFLICT 7 7 C -> G (IN REF. 1).  
 FT CONFLICT 85 85 S -> P (IN REF. 3 AND 5).  
 FT CONFLICT 97 97 Q -> R (IN REF. 3 AND 5).  
 FT CONFLICT 530 531 SL -> P (IN REF. 2).  
 FT CONFLICT 542 542 A -> P (IN REF. 3).  
 FT CONFLICT 548 549 MP -> TL (IN REF. 3).  
 FT CONFLICT 578 579 AQ -> PR (IN REF. 3).  
 FT CONFLICT 704 705 NQ -> KP (IN REF. 3).  
 FT CONFLICT 720 720 R -> Q (IN REF. 3).  
 FT CONFLICT 726 726 D -> N (IN REF. 3).  
 FT CONFLICT 843 846 TQPR -> HGSF (IN REF. 4).  
 FT CONFLICT 882 884 OGP -> SGS (IN REF. 4).  
 FT CONFLICT 894 894 P -> L (IN REF. 1).  
 FT CONFLICT 1031 1032 PP -> RQ (IN REF. 1 AND 4).  
 FT CONFLICT 1091 1091 D -> V (IN REF. 4).  
 FT CONFLICT 1124 1124 A -> R (IN REF. 4).  
 FT CONFLICT 1127 1133 EPGARNP -> GAGGIGT (IN REF. 4).  
 FT CONFLICT 1253 1253 A -> P (IN REF. 2).  
 FT CONFLICT 1257 1257 Q -> T (IN REF. 2).  
 FT CONFLICT 1352 1352 E -> R (IN REF. 4).  
 FT CONFLICT 1736 1736 AA: 171781 MW: 722A912568154B5F CRC64;  
 SQ SEQUENCE  
 Query Match 59.8%; Score 55; DB 1; Length 1736;  
 Best Local Similarity 66.7%; Pred. No. 5;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 PAGPWGPNRGKRGV 15  
 Db 1184 PRGPAGPNCADPGQ 1198  
 RESULT 21  
 YK61\_CAEEL STANDARD; PRT; 305 AA.  
 ID YK61\_CAEEL  
 AC P34340;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Putative cuticle collagen C29E4.1.  
 GN C29E4.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxId=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Faveello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Shownkeen R.,  
 RA Sims M., Smailon N., Smith A., Smith M., Sonhammer E., Staden R.,

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,  
 RA Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE NEM. FROM ITS ENVIRONMENT (BY SIMILARITY).  
 CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLETED WITHIN THE  
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINKS (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE  
 CC COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L23651; AAA27958.1; -.  
 DR PIR: S44767; S44767.  
 DR WormPep: C29E4.1; CE00083.  
 DR InterPro: IPR002486; COL\_cuticle\_N.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 2.  
 DR Pfam: PF01484; COL\_cuticle\_N; 1.  
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;  
 KM Multigene family; Collagen.  
 FT DOMAIN 96 125 TRIPLE-HELICAL REGION.  
 FT DOMAIN 142 204 TRIPLE-HELICAL REGION.  
 FT DOMAIN 208 252 TRIPLE-HELICAL REGION.  
 FT DOMAIN 256 270 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 305 AA; 30623 MW; B44CCE8D5501E27D CRC64;  
 Query Match 58.7%; Score 54; DB 1; Length 305;  
 Best Local Similarity 66.7%; Pred. No. 1.4;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 PAGPMPGNGKDKYVG 15  
 DB 140 PAGPMPGNGKDKYVG 154  
 RESULT 22  
 CA13\_BOVIN  
 ID CA13\_BOVIN STANDARD; PRT; 1049 AA.  
 AC P04258;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3A1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE OF 1-242.  
 RX MEDLINE=80026026; PubMed=488906;  
 RA Kuhn K.;  
 RA Fietzek P.P., Allmann H., Rautenberg J., Henkel W., Wachter E.,  
 RT "The covalent structure of calf skin type III collagen. I. The amino  
 RT acid sequence of the amino terminal region of the alpha 1(III) chain  
 RT (positions 1-222).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RN [2]  
 RP SEQUENCE OF 243-422.  
 RX MEDLINE=80026027; PubMed=488907;

RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2  
 RT (positions 223-402).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RN [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The  
 RT amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4  
 RT (positions 403-551).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RN [4]  
 RP SEQUENCE OF 572-808.  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Glanville R.W., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB5  
 RT (positions 552-788).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RN [5]  
 RP SEQUENCE OF 809-947.  
 RX MEDLINE=80026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino  
 RT acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A  
 RT (position 789-927).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RN [6]  
 RP SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Allmann H., Fietzek P.P., Glanville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino  
 RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha  
 RT 1(III)CB9B (positions 928-1028).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC ALONG WITH TYPE I COLLAGEN.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLISINES.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 DR PIR: A02862; CGB075.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001007; VWFC.  
 DR Pfam: PF01391; Collagen; 17.  
 DR PROSITE: PS01208; VWFC; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KM Glycoprotein; Collagen.  
 FT DOMAIN 1 14 NONHELICAL REGION (N-TERMINAL).  
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1041 1049 NONHELICAL REGION (C-TERMINAL).  
 FT MOD\_RES 95 95 HYDROXYLATION.  
 FT MOD\_RES 107 107 HYDROXYLATION.  
 FT MOD\_RES 119 119 HYDROXYLATION.  
 FT MOD\_RES 938 938 HYDROXYLATION.  
 FT MOD\_RES 950 950 HYDROXYLATION.  
 FT CARBOHYD 107 107 O-LINKED (GAL. . .).  
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).  
 FT DISULFID 1040 1040 INTERCHAIN.  
 FT DISULFID 1041 1041 INTERCHAIN.  
 SQ SEQUENCE 1049 AA; 93651 MW; BEEC33D1C66EC9A3 CRC64;  
 Query Match 58.7%; Score 54; DB 1; Length 1049;  
 Best Local Similarity 66.7%; Pred. No. 4.3;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PAGPMPGNGKDKYVG 15  
 DB 985 PVGPMPGNGKDKYVG 999

```

RESULT 23
CA21_RANCA STANDARD: PRT: 1355 AA.
ID CA21_RANCA
AC 042350:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(I) chain precursor.
GN COL1A2.
OS Rana catesbeiana (Bull frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranioidea; Ranidae; Rana.
ON NCBI_Taxid=8400;
RX NCBI_Taxid=8400;
RC TISSUE-Tail;
RP SEQUENCE FROM N.A.
RX MEDLINE=97417499; PubMed=9272872;
RA Asahina K., Oofusa K., Obara M., Yoshizato K.;
RT "Cloning and characterization of the full length cDNA encoding alpha2
RT type I collagen of bullfrog Rana catesbeiana.";
RL Gene 194:283-289(1997).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
CC (FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
CC HYDROXYAPATITE.
CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sb.ch/announce/
CC or send an email to license@isb-sb.ch).
CC -----
DR EMBL: D88764; BAA22380.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR Pfam: PF01410; COLFI.1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI.1.
DR KX Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KV Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT PROPEP 23 82 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT CHAIN 83 1093 COLLAGEN ALPHA 2(I) CHAIN.
FT PROPEP 1094 1355 CARBOXYL-TERMINAL PROPEPTIDE
(BY SIMILARITY).
FT CARBOHD 1206 1206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 1256 1256 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1355 AA; 127643 MW; CB793AD5D6F4D12A CRC64;

Query Match: 58.7%; Score 54; DB 1; Length 1355;
Best Local Similarity: 60.0%; Pred. No. 5.5;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

```

DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_Taxid=9606;
RX NCBI_Taxid=9606;
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Koko L., Kontusari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516(1989).
RN [12]
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742(1989).
RN [13]
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=557335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanoogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164(1977).
RN [14]
RP REVISIONS.
RA Seyer J.M.;
RT Submitted (DEC-1977) to the PIR data bank.
RN [15]
RP SEQUENCE OF 399-727.
RX MEDLINE=7900343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBR peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411(1978).
RN [16]
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
RN [17]
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalgleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337(1988).
RN [18]
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
RN [19]
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenblom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
RN [110]
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;

```





RP SEQUENCE FROM N.A. (SHORT ISOFORM).  
 RC STRAIN-BALB/C; TISSUE=Liver;  
 RX MEDLINE=94245707; PubMed=8188673;  
 RA Rehn M.V., Hintikka E., Pihlajaniemi T.;  
 RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,  
 RT partial structure of the corresponding gene, and comparison of the  
 RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen  
 RT chain.";  
 RL J. Biol. Chem. 269:13929-13935(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).  
 RA Rehn M., Hintikka E., Pihlajaniemi T.;  
 RT "Characterization of the mouse gene for the alpha-1 chain of type  
 RT XVIII collagen (COL18A1) reveals that the three variant N-terminal  
 RT polypeptide forms are transcribed from two widely separated  
 RT promoters.";  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).  
 RX MEDLINE=94240112; PubMed=8183894;  
 RA Rehn M.V., Pihlajaniemi T.;  
 RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the  
 RT collagenous sequence, a distinct tissue distribution, and homology  
 RT with type XV collagen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
 RN [4]  
 RP SEQUENCE OF 240-1527 FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=94240111; PubMed=8183893;  
 RA Oh S.P., Kamagata Y., Murgaki Y., Timmons S., Ooshima A., Olsen B.R.;  
 RT "Isolation and sequencing of cDNAs for proteins with multiple domains  
 RT of Gly-Xaa-Yaa repeats identify a distinct family of collagenous  
 RT proteins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
 RN [5]  
 RP CHARACTERIZATION OF ENDOSTATIN AND PARTIAL SEQUENCE.  
 RX MEDLINE=97160848; PubMed=9008168;  
 RA O'Reilly M.S., Boehm T., Snider Y., Fukai N., Vasios G., Lane W.S.,  
 RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
 RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor  
 RT growth.";  
 RL Cell 88:277-285(1997).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.  
 RX MEDLINE=98169382; PubMed=9501087;  
 RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;  
 RT "Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A  
 RT resolution.";  
 RL EMBO J. 17:1656-1664(1998).  
 CC -I- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
 CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
 CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
 CC FACTOR SIGNALING.  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- PPM: PROLINS ARE AT THE THIRD POSITION OF THE TRIPLETT REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -I- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; U34609; AAC52901.1; JOINED.  
 DR EMBL; U34610; AAC52901.1; JOINED.  
 DR EMBL; U34611; AAC52901.1; JOINED.  
 DR EMBL; U34612; AAC52901.1; JOINED.  
 DR EMBL; U34613; AAC52901.1; JOINED.  
 DR EMBL; U03716; AAC52901.1; JOINED.  
 DR EMBL; U03718; AAC52901.1; JOINED.  
 DR EMBL; U03715; AAC52902.1; JOINED.  
 DR EMBL; U34607; AAC52902.1; JOINED.  
 DR EMBL; U34608; AAC52902.1; JOINED.  
 DR EMBL; U34609; AAC52902.1; JOINED.  
 DR EMBL; U34610; AAC52902.1; JOINED.  
 DR EMBL; U34611; AAC52902.1; JOINED.  
 DR EMBL; U34612; AAC52902.1; JOINED.  
 DR EMBL; U34613; AAC52902.1; JOINED.  
 DR EMBL; U03716; AAC52902.1; JOINED.  
 DR EMBL; U03718; AAC52902.1; JOINED.  
 DR EMBL; U11636; AAC52178.1; JOINED.  
 DR EMBL; L22545; AAC19787.1; JOINED.  
 DR PDB; 1KOE; 16-FEB-99.  
 DR MGD; MGI:88451; COL18A1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR001791; Laminin\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam; PF01391; Collagen; 8.  
 DR Pfam; PF02210; TSPN; 2.  
 DR SMART; SM00282; Lamc; 1.  
 DR SMART; SM00210; TSPN; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
 KW 3D-structure.  
 FT SIGNAL 1 26  
 FT CHAIN 27 1527  
 FT CHAIN 1344 1527  
 FT DOMAIN 27 538  
 FT DOMAIN 539 565  
 FT DOMAIN 566 575  
 FT DOMAIN 576 649  
 FT DOMAIN 650 673  
 FT DOMAIN 674 795  
 FT DOMAIN 796 818  
 FT DOMAIN 819 901  
 FT DOMAIN 902 915  
 FT DOMAIN 916 957  
 FT DOMAIN 958 970  
 FT DOMAIN 971 1043  
 FT DOMAIN 1044 1053  
 FT DOMAIN 1054 1086  
 FT DOMAIN 1087 1098  
 FT DOMAIN 1099 1122  
 FT DOMAIN 1123 1129  
 FT DOMAIN 1130 1181  
 FT DOMAIN 1182 1194  
 FT DOMAIN 1195 1212  
 FT DOMAIN 1213 1527  
 FT CARBOHYD 338 338  
 FT CARBOHYD 700 700  
 FT DISULFID 1376 1516  
 FT DISULFID 1478 1508  
 FT SITE 1104 1106  
 FT VASSPLIC 1 212  
 FT VASSPLIC 213 238  
 FT CONFLICT 900 900  
 FT CONFLICT 947 947  
 FT CONFLICT 964 964  
 FT CONFLICT 1157 1157  
 FT CONFLICT 1266 1266  
 FT CONFLICT 1276 1276  
 FT CONFLICT 1437 1437  
 SQ SEQUENCE 1527 AA; 156008 MW; 9645045AFL40B513 CRC64;

Query Match

58.7%; Score 54; DB 1; Length 1527;

CELL ATTACHMENT SITE (POTENTIAL).  
 MISSING (IN SHORT ISOFORM).  
 AVPTQLPFGSNLADPLRSPAPDF -> MAPRWHLDLVL  
 TSLVLLVAVRSMAE (IN SHORT ISOFORM).  
 P -> L (IN REF. 4).  
 P -> F (IN REF. 4).  
 A -> R (IN REF. 4).  
 R -> P (IN REF. 4).  
 P -> L (IN REF. 4).  
 L -> F (IN REF. 4).  
 L -> V (IN REF. 4).  
 MW; 9645045AFL40B513 CRC64;



Best Local Similarity	66.7%	Pred. No. 6.1	Matches 10;	conservative 0;	mismatches 5;	Indels 0;	Gaps 0;
OY	1	PAGMGPNKGDKGV 15					
DB	580	PPGPGGPPKDGDPG 594					
RESULT 26							
CA15_HUMAN	STANDARD:	PRT, 1838 AA.					
AC	P20908:						
DT	01-FEB-1991 (Rel. 17, Created)						
DT	01-MAY-1992 (Rel. 22, Last sequence update)						
DT	16-OCT-2001 (Rel. 40, Last annotation update)						
DE	Collagen alpha 1(V) chain precursor.						
GN	COL5A1..						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.						
OX	NCBI_TaxID:9606;						
RA	SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.						
RP	MEDLINE=91302336; PubMed=2071595;						
RA	Takahara K., Seto Y., Okamoto N., Noda A., Yaei Y.,						
RA	Kato I.;						
RT	"Complete primary structure of human collagen alpha 1 (V) chain.";						
RL	J. Biol. Chem. 266:13124-13129(1991).						
RL	[2]						
RP	SEQUENCE OF 621-822.						
RC	TISSUE=Chorioamniotic membrane;						
RC	MEDLINE=89227189; PubMed=2496661;						
RA	Seyer J.M., Kang A.H.;						
RT	"Covalent structure of collagen: amino acid sequence of three						
RT	cyanogen bromide-derived peptides from human alpha 1(V) collagen						
RT	chain.";						
RL	Arch. Biochem. Biophys. 271:120-129(1989).						
RL	[3]						
RP	SEQUENCE OF 823-950, AND HEPARIN-BINDING.						
RA	MEDLINE=90366601; PubMed=2203476;						
RA	Yaei Y., Hashimoto K., Koltashli H., Takahara K., Ito M., Kato I.;						
RT	"Primary structure of the heparin-binding site of type V collagen.";						
RL	Biochim. Biophys. Acta 1035:139-145(1990).						
RL	[4]						
RP	SEQUENCE OF 556-571.						
RP	TISSUE=Placenta;						
RC	MEDLINE=92239022; PubMed=1571108;						
RA	Mann K.;						
RT	"Isolation of the alpha 3-chain of human type V collagen and						
RT	characterization by partial sequencing.";						
RL	Biol. Chem. Hoppe-Seyler 373:69-75(1992).						
RL	[5]						
RP	SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.						
RC	TISSUE=Chorioamniotic membrane;						
RC	MEDLINE=94237164; PubMed=818182;						
RA	Moradi-Ameli M., Rousseau J.C., Klemen J.P., Champliand M.F.,						
RA	Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;						
RT	"Diversity in the processing events at the N-terminus of type-V						
RT	collagen.";						
RL	Eur. J. Biochem. 221:987-995(1994).						
RL	[6]						
RP	VARIANT EDS1 SER-1639.						
RA	MEDLINE=97195540; PubMed=9042913;						
RA	de Paeye A., Nuytinc L., Hauser I., Anton-Lamprecht I.,						
RT	Naeyaert J.-M.;						
RT	"Mutations in the COL5A1 gene are causal in the Ehlers-Danlos						
RT	syndromes I and II.";						
RT	Am. J. Hum. Genet. 60:547-554(1997).						
CC	-1- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN						
CC	(FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE						

[illegible]

```

FT MOD_RES 774 774 HYDROXYLATION.
FT MOD_RES 780 780 HYDROXYLATION.
FT MOD_RES 789 789 HYDROXYLATION.
FT MOD_RES 795 795 HYDROXYLATION.
FT MOD_RES 804 804 HYDROXYLATION.
FT MOD_RES 807 807 HYDROXYLATION.
FT MOD_RES 810 810 HYDROXYLATION.
FT MOD_RES 816 816 HYDROXYLATION.
FT MOD_RES 819 819 HYDROXYLATION.
FT MOD_RES 834 834 HYDROXYLATION.
FT MOD_RES 846 846 HYDROXYLATION.
FT MOD_RES 861 861 HYDROXYLATION.
FT MOD_RES 864 864 HYDROXYLATION.
FT MOD_RES 870 870 HYDROXYLATION.
FT MOD_RES 873 873 HYDROXYLATION.
FT MOD_RES 876 876 HYDROXYLATION.
FT MOD_RES 882 882 HYDROXYLATION.
FT MOD_RES 888 888 HYDROXYLATION.
FT MOD_RES 891 891 HYDROXYLATION.
FT MOD_RES 897 897 HYDROXYLATION.
FT MOD_RES 903 903 HYDROXYLATION.
FT MOD_RES 906 906 HYDROXYLATION.
FT MOD_RES 930 930 HYDROXYLATION.
FT MOD_RES 945 945 HYDROXYLATION.
FT MOD_RES 1017 1017 HYDROXYLATION.
FT MOD_RES 1020 1020 HYDROXYLATION.
FT MOD_RES 1023 1023 HYDROXYLATION.
FT MOD_RES 1029 1029 HYDROXYLATION.
FT MOD_RES 1221 1221 HYDROXYLATION.
FT MOD_RES 1224 1224 HYDROXYLATION.
FT MOD_RES 1467 1467 HYDROXYLATION.
FT MOD_RES 1470 1470 HYDROXYLATION.
FT VARIANT 1639 1639 /FTID=VAR.001808.
FT CONFLICT 641 641 E -> G (IN REF. 2).
FT CONFLICT 650 650 R -> L (IN REF. 2).
FT CONFLICT 663 663 R -> E (IN REF. 2).
FT CONFLICT 668 668 E -> Q (IN REF. 2).
FT CONFLICT 677 677 K -> Q (IN REF. 2).
FT CONFLICT 684 684 L -> P (IN REF. 2).
FT CONFLICT 692 699 PPGPGV -> VTGPGAP (IN REF. 2).
FT CONFLICT 727 727 G -> Q (IN REF. 2).
FT CONFLICT 741 741 P -> L (IN REF. 2).
FT CONFLICT 747 747 L -> Q (IN REF. 2).
FT CONFLICT 753 753 P -> A (IN REF. 2).
FT CONFLICT 759 759 D -> N (IN REF. 2).
FT CONFLICT 776 777 GO -> OK (IN REF. 2).
FT CONFLICT 849 855 GGPNGDP -> IGPGPR (IN REF. 3).
FT CONFLICT 894 894 N -> D (IN REF. 3).
SQ SEQUENCE 1838 AA; 183616 MW; 7D58239C0D77BD4E CRC64;

Query Match          58.7%; Score 54; DB 1; Length 1838;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94131298; PubMed=8299960;
RA 'Levy A.D., Kramer J.M.;
RT "Identification, sequence and expression patterns of the
RT Caenorhabditis elegans col-36 and col-40 collagen-encoding genes.";
RN Gene 137:281-285(1993).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RA Pauley A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -! FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -! SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -! SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
CC COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L15418; AAA17445.1; -.
DR EMBL; U14635; AAK84492.1; -.
DR WormPep; C27H5.5; CE06893.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen_3.
DR Pfam; PF01484; Col_cuticle_N; 1.
FT CUTICLE; Connective tissue; Repeat; Multigene family; Collagen.
FT DOMAIN 89 105 TRIPLE-HELICAL REGION.
FT DOMAIN 118 150 TRIPLE-HELICAL REGION.
FT DOMAIN 167 187 TRIPLE-HELICAL REGION.
FT DOMAIN 194 226 TRIPLE-HELICAL REGION.
FT DOMAIN 231 257 TRIPLE-HELICAL REGION.
FT DOMAIN 260 295 TRIPLE-HELICAL REGION.
SQ SEQUENCE 307 AA; 30126 MW; 9C7206CC18660F0B CRC64;

Query Match          57.6%; Score 53; DB 1; Length 307;
Best Local Similarity 66.7%; Pred. No. 1.9;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

RT *Sequence comparisons of developmentally regulated collagen genes of
RT Caenorhabditis elegans.*
RL Gene 76:331-344(1989).
CC
CC -1- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
CC -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLETED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS.
CC -1- SIMILARITY: TO OTHER COLLAGENS. STRONG. TO OTHER CUTICLE
CC COLLAGENS. COL-2 AND COL-6 BELONGS TO THE SAME GROUP OF COLLAGEN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M25477; AAA27991.1; -.
CC PIR: J50167; J50167.
CC DR InterPro: IPR002486; Col_cuticle_N.
CC DR InterPro: IPR000087; Collagen.
CC DR Pfam: PF01391; Collagen; 2.
CC DR Pfam: PF01484; Col_cuticle_N; 1.
CC KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.
CC FT DOMAIN 95 115 GLY-RICH
CC FT DOMAIN 142 171 TRIPLE-HELICAL REGION.
CC FT DOMAIN 189 212 TRIPLE-HELICAL REGION.
CC FT DOMAIN 216 248 TRIPLE-HELICAL REGION.
CC FT DOMAIN 253 279 TRIPLE-HELICAL REGION.
CC FT DOMAIN 282 320 TRIPLE-HELICAL REGION.
CC SQ SEQUENCE 329 AA; 32561 MW; 52C295E77E76418 CRC64;

Query Match 57.6%; Score 53; DB 1; Length 329;
Best Local Similarity 66.7%; Pred. No. 2;
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMGPNKGDKGVG 15
Db 146 PGPEGPNGDKGKG 160

RESULT 29
PSPD_MOUSE
ID PSPD_MOUSE STANDARD: PRT; 374 AA.
AC P50404;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
DE SFTPD OR SFTPA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBA; TISSUE=Lung;
RX MEDLINE=96094460; PubMed=7498852;
RA Motwani M., White R.A., Guo N., Dowler L.L., Tauber A.I., Sastri K.N.;
RT "Mouse surfactant protein-D. cDNA cloning, characterization, and gene
RT localization to chromosome 14."
RL J. Immunol. 155:5671-5677(1995).
CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,

```

```

CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: I40156; AAA92021.1; -.
CC HSSP: P35247; 1B08.
CC MGD: MGI:109515; Sftpd.
CC DR InterPro: IPR000087; Collagen.
CC DR InterPro: IPR001304; Lectin_c.
CC DR Pfam: PF01391; Collagen; 3.
CC DR Pfam: PF00059; Lectin_c; 1.
CC DR SMART: SM00034; CLECT; 1.
CC DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
CC DR PROSITE: PS00411; C-TYPE-LECTIN_2; 1.
CC KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
CC Signal; Lectin; Collagen; Repeat; Coiled coil.
CC FT SIGNAL 1 19 BY SIMILARITY.
CC FT CHAIN 20 374 PULMONARY SURFACTANT-ASSOCIATED PROTEIN
CC FT DOMAIN 45 221 D.
CC FT DOMAIN 222 253 COLLAGEN-LIKE.
CC FT DOMAIN 278 374 COILED COIL (POTENTIAL).
CC FT DISULFID 280 372 C-TYPE LECTIN (SHORT FORM).
CC FT DISULFID 350 364 BY SIMILARITY.
CC FT CARBOHYD 89 89 N-LINKED (GLCNAC... (POTENTIAL).
CC SQ SEQUENCE 374 AA; 37688 MW; FE034261263F43E4 CRC64;

Query Match 57.6%; Score 53; DB 1; Length 374;
Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKGDKGVG 15
Db 115 PAGKEGPGSGKGNIG 129

RESULT 30
PSPD_RAT
ID PSPD_RAT STANDARD: PRT; 374 AA.
AC P35248;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)
DE (CP4).
OS Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.
RC TISSUE=Lung;
RX MEDLINE=92112913; PubMed=1370483;
RA Shimizu H., Fisher J.H., Pepst F., Benson B., Lau K., Mason R.J.,
RA Voelker D.R.;
RT "Primary structure of rat pulmonary surfactant protein D. cDNA and
RT deduced amino acid sequence."
RL J. Biol. Chem. 267:1853-1857(1992).
RN [2]
RP SEQUENCE OF 73-95 AND 153-180.
RC TISSUE=Lung;
RX MEDLINE=90001186; PubMed=2675969;
RA Persson A., Chang D., Rust K., Moxley M., Longmore W., Crouch E.;

```

```

RT      "Purification and biochemical characterization of CP4 (SP-D), a
RL      collagenous surfactant-associated protein.";
CC      Biochemistry 28:6361-6367(1989).
CC      -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
CC      MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
CC      EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
CC      EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC      -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC      -1- SUBCELLULAR LOCATION: Extracellular.
CC      -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC      PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC      CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC      HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC      -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC      -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; M81231; AAA42170.1; -.
DR      PIR; A42046; A42046.
DR      HSSP; P35247; 1B08.
DR      InterPro; IPR000087; Collagen.
DR      InterPro; IPR001304; lectin_c.
DR      Pfam; PF01391; Collagen; 3.
DR      Pfam; PF00059; lectin_c; 1.
DR      SMART; SM00034; CLECT; 1.
DR      PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR      PROSITE; PS50041; C-TYPE LECTIN_2; 1.
KW      Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
KW      Signal; Lectin; Collagen; Repeat; Coiled coil.
FT      SIGNAL      1      19
FT      CHAIN       20      374
FT      FT
FT      DOMAIN      45      221      D.
FT      DOMAIN      222      253      COLLAGEN-LIKE.
FT      DOMAIN      278      374      COILED COIL (POTENTIAL).
FT      DISULFID     280      372      C-TYPE LECTIN (SHORT FORM).
FT      DISULFID     350      364      BY SIMILARITY.
FT      CARBOHYD     89      89      N-LINKED (GLCNAC. . .).
FT      MOD_RES      77      77      HYDROXYLATION.
FT      MOD_RES      86      86      HYDROXYLATION.
FT      MOD_RES      95      95      HYDROXYLATION.
FT      MOD_RES      98      98      HYDROXYLATION.
FT      MOD_RES      170      170      HYDROXYLATION.
FT      MOD_RES      176      176      HYDROXYLATION.
FT      CONFLICT     89      89      N -> E (IN REF. 2).
FT      CONFLICT     164      164      K -> C (IN REF. 2).
SO      SEQUENCE    374 AA; 37561 MW; DB2B5E399DB4A3C CRC64;

```

```

Query Match      57.6%; Score 53; DB 1; Length 374;
Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

```

OY      1 PAGPWGPNKGDKVG 15
      111 11:11 1:1
DB      115 PAGKEGPGSGKGNIG 129

```

Search completed: November 1, 2002, 12:53:54  
 Job time : 8 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:19:21 ; Search time 21.5 Seconds

(without alignments)  
120.694 Million cell updates/sec

Title: US-09-529-691A-3

Sequence: 1 PAGPMGPNKGDKVKG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	75.0	460	5	09GZF7
2	66	71.7	291	5	019470
3	63	68.5	1366	4	015177
4	62	67.4	326	2	09AGC6
5	61	66.3	410	2	09F691
6	60	65.2	358	12	09J304
7	60	65.2	356	11	091VL4
8	59	64.1	380	2	09F688
9	59	64.1	380	2	09F688
10	59	64.1	739	11	070575
11	59	64.1	751	11	09R1N9
12	59	64.1	1017	11	099K41
13	58	63.0	434	2	09ACN4
14	57	62.0	1051	5	026055
15	57	62.0	1352	13	090YJ0
16	57	62.0	1419	11	063123

17	57	62.0	1442	11	062031	062031 mus musculus
18	57	62.0	1442	11	062033	062033 mus musculus
19	57	62.0	1459	11	062032	062032 mus musculus
20	56	60.9	303	5	093208	093208 caenorhabd
21	56	60.9	588	12	09QEK6	09QEK6 cynomolgus
22	56	60.9	888	13	090796	090796 gallus gall
23	56	60.9	922	5	044367	044367 mytilus edu
24	56	60.9	1076	5	026154	026154 plasmodium
25	56	60.9	1418	13	09W7R9	09W7R9 cynops pyrr
26	56	60.9	1420	13	09W377	09W377 gallus gall
27	56	60.9	1486	13	091717	091717 xenopus lae
28	56	60.9	1491	13	091718	091718 xenopus lae
29	56	60.9	1767	4	09UIT4	09UIT4 homo sapien
30	56	60.9	1806	4	09UIT5	09UIT5 homo sapien
31	56	60.9	1818	4	09UIT6	09UIT6 homo sapien
32	55	59.8	172	5	000487	000487 hydra sp. m
33	55	59.8	196	2	09ACM9	09ACM9 streptococ
34	55	59.8	207	4	014044	014044 homo sapien
35	55	59.8	215	13	093392	093392 coturnix co
36	55	59.8	284	5	025582	025582 teladorsagi
37	55	59.8	284	5	025581	025581 teladorsagi
38	55	59.8	301	5	022379	022379 caenorhabd
39	55	59.8	301	5	022380	022380 caenorhabd
40	55	59.8	347	6	09XT25	09XT25 bos taurus
41	55	59.8	363	5	022326	022326 caenorhabd
42	55	59.8	464	13	090412	090412 brachydanio
43	55	59.8	678	13	093486	093486 oncorhynch
44	55	59.8	890	5	077087	077087 alvinella p
45	55	59.8	902	5	016161	016161 mytilus edu

#### ALIGNMENTS

RESULT 1  
ID 09GZF7 PRELIMINARY; PRT; 460 AA.  
AC 09GZF7;  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 41.0 KDA PROTEIN.  
GN C18H7.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Tin-Woliam A.; Fironick W.;  
RT "The sequence of C. elegans cosmid C18H7.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submissions.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF067507; AAF98609.1; -;  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen\_2.  
DR Pfam: PF01484; COL\_cuticle\_N; 1.  
KW Hypothetical protein.

5Q SEQUENCE 460 AA; 41016 MW; BF99CC80770E2202 CRC64;

Query Match 75.0%; Score 69; DB 5; Length 460;  
Best Local Similarity 80.0%; Pred. No. 0.021;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 237 PAGPMPNGKDGKVG 251

RESULT 2

019470 ID 019470 PRELIMINARY; PRT; 291 AA.

AC 019470;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE F15A2.1 PROTEIN.  
GN F15A2.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RN [1]  
RP SEQUENCE FROM N.A.

RA Gregory J.;  
RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology."  
RL Science 282:2012-2018(1998).  
DR EMBL: 270207; CA94128.1; -.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR002486; Col\_cuticle\_N.  
DR Pfam; PF01391; Collagen; 2.  
DR Pfam; PF01484; Col\_cuticle\_N; 1.  
SQ SEQUENCE 291 AA; 28482 MW; BF6AB31C7D5E2CA4 CRC64;

Query Match 71.7%; Score 66; DB 5; Length 291;  
Best Local Similarity 80.0%; Pred. No. 0.036;  
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 183 PAGPMPNGKDGKVG 197

RESULT 3

015177 ID 015177 PRELIMINARY; PRT; 1366 AA.

AC 015177;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PREPRO-ALPHA2(I) COLLAGEN PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 1-765 FROM N.A.  
RX MEDLINE=88339824; PubMed=3421913;  
RA Kuitavanem H., Tromp G., Chu M., Prockop D.J.;  
RT "Structure of a full-length cDNA clone for the preproalpha2(I) chain  
of human type I procollagen."  
RL Biochem. J. 252:633-640(1988).  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE=88058962; PubMed=2824475;

RA de Wet W., Bernard M., Benson-Chanda V., Chu M., Dickson L., Well D.,  
RA Ramirez F.;  
RT "Organization of the human pro-alpha2(I) collagen gene."  
RL J. Biol. Chem. 262:16032-16036(1987).  
RN [3]  
RP SEQUENCE OF 959-1351 FROM N.A.  
RX MEDLINE=90304220; PubMed=2364107;  
RA Makela J.K., Vuorio T., Vuorio E.;  
RT "Growth-dependent modulation of type I collagen production and mRNA  
levels in cultured human skin fibroblasts."  
RL Biochim. Biophys. Acta 1049:171-176(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Dalgleish R.;  
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.

RX MEDLINE=97169389; PubMed=9016532;

RA Dalgleish R.;

RT "The human type I collagen mutation database."

RL Nucleic Acids Res. 25:181-187(1997).

DR EMBL: 274616; CA98969.1; -.  
DR InterPro; IPR000087; Collagen.

DR InterPro; IPR000885; Fib\_collagen\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR Prodom; PD002078; Fib\_collagen\_C; 1.

DR SMART; SM00038; COLFI; 1.

KW Signal; Collagen.

FT SIGNAL 1 22 POTENTIAL.

SQ SEQUENCE 1366 AA; 129337 MW; 5796859B6E50286C CRC64;

Query Match 68.5%; Score 63; DB 4; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.52;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
|||||  
DB 1055 PAGPMPNGKDGRTG 1069

RESULT 4

09AGC6 ID 09AGC6 PRELIMINARY; PRT; 326 AA.

AC 09AGC6;

DT 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE COLLAGEN-LIKE PROTEIN B (FRAGMENT).

OS Streptococcus pyogenes.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1314;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=AP12;

RA Rasmussen M., Bjorck L.;

RT "Phase-variation of a novel collagen-like surface protein in  
Streptococcus pyogenes."

RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AF336811; AAK30077.1; -.  
DR InterPro; IPR000087; Collagen.

DR Pfam; PF01391; Collagen; 3.

FT NON\_TER 1 326

FT NON\_TER 1 326

SQ SEQUENCE 326 AA; 33383 MW; BBE52296FF24245B CRC64;

Query Match 67.4%; Score 62; DB 2; Length 326;  
Best Local Similarity 73.3%; Pred. No. 0.16;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
|||||

Db 126 PAGPVGPKDGEAG 140

## RESULT 5

Q9F691 ID Q9F691 PRELIMINARY: PRT: 410 AA.  
AC Q9F691;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).  
OS Streptococcus pyogenes.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1314;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AP12;  
RX MEDLINE=20490571; PubMed=11035747;  
RA Rasmussen M., Eden A., Bjorck L.;  
RT "ScLA, a novel collagen-like surface protein of streptococcus  
pyogenes.";  
RL Infect. Immun. 68:6370-6377(2000).  
DR EMBL: AF296332; AAC30212.1; -;  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen: 2;  
DR PRINTS: PR01217; PRICHEXTENSN.  
FT NON\_TER 1 1  
FT NON\_TER 410 410  
SQ SEQUENCE 410 AA; 42464 MW; D4559670E7ECDADF CRC64;

Query Match 66.3%; Score 61; DB 2; Length 410;  
Best Local Similarity 73.3%; Pred. No. 0.29;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPVGPKDGEAG 15  
Db 118 PAGPVGPKDGEAG 132

## RESULT 6

Q9J3U4 ID Q9J3U4 PRELIMINARY: PRT: 358 AA.  
AC Q9J3U4;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE COLLAGEN-LIKE PROTEIN 1.  
OS Ectocarpus siliculosus virus.  
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae.  
OX NCBI\_TaxID=37665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ESV-1;  
RX MEDLINE=20192171; PubMed=10725207;  
RA Delaroque N., Wolf S., Muller D.G., Knippers R.;  
RT "Characterization and immunolocalization of major structural proteins  
in the brown algal virus ESV-1.";  
RL Virology 269:148-155(2000).  
DR EMBL: AF204951; AAF28321.1; -;  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen: 3;  
SQ SEQUENCE 358 AA; 33690 MW; A9A07CC077AA025A CRC64;

Query Match 65.2%; Score 60; DB 12; Length 358;  
Best Local Similarity 73.3%; Pred. No. 0.35;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGPVGPKDGEAG 15  
Db 264 PAGPVGPKDGEAG 278

## RESULT 7

Q91VLA ID Q91VLA PRELIMINARY: PRT: 566 AA.  
AC Q91VLA;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE UNKNOWN (PROTEIN FOR IMAGE:3481648) (FRAGMENT).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST TUMOR;  
RA Strausberg R.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC012438; AAH12438.1; -;  
FT NON\_TER 1 1  
SQ SEQUENCE 566 AA; 56663 MW; 2E845D0DFC7BE0D5 CRC64;

Query Match 65.2%; Score 60; DB 11; Length 566;  
Best Local Similarity 73.3%; Pred. No. 0.57;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 PAGPVGPKDGEAG 15  
Db 255 PAGPVGPKDGEAG 269

## RESULT 8

Q9U336 ID Q9U336 PRELIMINARY: PRT: 328 AA.  
AC Q9U336;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE W07A12.5 PROTEIN.  
GN W07A12.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;  
OC Rhabditidae; Pelodermidae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Thomas K.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C. elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL: Z68320; CAB63328.1; -;  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen: 2;  
SQ SEQUENCE 328 AA; 33720 MW; 6178EEF08F884AC5 CRC64;

Query Match 64.1%; Score 59; DB 5; Length 328;  
Best Local Similarity 66.7%; Pred. No. 0.45;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPVGPKDGEAG 15  
Db 221 PAGPVGPKDGEAG 235

## RESULT 9

Q9F688 ID Q9F688 PRELIMINARY: PRT: 380 AA.  
AC Q9F688;

DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxId=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=KT19;  
 RC MEDLINE=20490571; PubMed=11035747;  
 RA Rasmussen M., Eden A., Bjorck U.,  
 RT "ScLA, a novel collagen-like surface protein of streptococcus  
 pyogenes.";  
 RT Infect. Immun. 68:6370-6377(2000).  
 DR EMBL: AF296335; AAG30215.1;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 3.  
 FT NON\_TER 1  
 FT NON\_TER 380  
 SQ SEQUENCE 380 AA; 39224 MW; B6D283C3C7B7A17A4 CRC64;

Query Match 64.1%; Score 59; DB 2; Length 380;  
 Best Local Similarity 66.7%; Pred. No. 0.53;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDGKVG 15  
 DB 142 PGSPGPKGDKGKAG 156

RESULT 10  
 ID 070575 PRELIMINARY; PRT; 739 AA.  
 AC 070575  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE COLLAGEN TYPE XIII ALPHA-1 CHAIN.  
 GN COL13A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=GUT;  
 RX MEDLINE=98288296; PubMed=9624150;  
 RA Haeg P., Rehn M., Huhtala P., Vaisanen T., Tammunen M.,  
 RA Pihlajaniemi T.;  
 RT "Type XIII collagen is identified as a plasma membrane protein.";  
 RT J. Biol. Chem. 273:15590-15597(1998).  
 DR EMBL: U30292; AAC24314.1;  
 DR MGI: 1277201; COL13a1.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 9.  
 SQ SEQUENCE 739 AA; 72110 MW; F13951061381F017 CRC64;

Query Match 64.1%; Score 59; DB 11; Length 739;  
 Best Local Similarity 73.3%; Pred. No. 1.1;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDGKVG 15  
 DB 514 PAGPGPKGDKGPPG 528

RESULT 11  
 OY 09R1N9 PRELIMINARY; PRT; 751 AA.  
 AC 09R1N9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TYPE XIII COLLAGEN.  
 GN COL13A1 OR COL13A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99357014; PubMed=10429945;  
 RA Kvist A.P., Latvalento A., Sund M., Horelli-Kuitunen N., Rehn M.,  
 RA Palotie A., Beter D., Pihlajaniemi T.;  
 RT "Complete exon-intron organization and chromosomal location of the  
 RT gene for mouse type XIII collagen (coll13a1) and comparison with its  
 RT human homologue.";  
 RT Matrix Biol. 18:261-274(1999).  
 DR EMBL: AF063693; AAD50327.1;  
 DR EMBL: AF063666; AAD50327.1; JOINED.  
 DR EMBL: AF063667; AAD50327.1; JOINED.  
 DR EMBL: AF063668; AAD50327.1; JOINED.  
 DR EMBL: AF063669; AAD50327.1; JOINED.  
 DR EMBL: AF063670; AAD50327.1; JOINED.  
 DR EMBL: AF063671; AAD50327.1; JOINED.  
 DR EMBL: AF063672; AAD50327.1; JOINED.  
 DR EMBL: AF063673; AAD50327.1; JOINED.  
 DR EMBL: AF063674; AAD50327.1; JOINED.  
 DR EMBL: AF063675; AAD50327.1; JOINED.  
 DR EMBL: AF063676; AAD50327.1; JOINED.  
 DR EMBL: AF063677; AAD50327.1; JOINED.  
 DR EMBL: AF063678; AAD50327.1; JOINED.  
 DR EMBL: AF063679; AAD50327.1; JOINED.  
 DR EMBL: AF063680; AAD50327.1; JOINED.  
 DR EMBL: AF063681; AAD50327.1; JOINED.  
 DR EMBL: AF063682; AAD50327.1; JOINED.  
 DR EMBL: AF063683; AAD50327.1; JOINED.  
 DR EMBL: AF063684; AAD50327.1; JOINED.  
 DR EMBL: AF063685; AAD50327.1; JOINED.  
 DR EMBL: AF063686; AAD50327.1; JOINED.  
 DR EMBL: AF063687; AAD50327.1; JOINED.  
 DR EMBL: AF063688; AAD50327.1; JOINED.  
 DR EMBL: AF063689; AAD50327.1; JOINED.  
 DR EMBL: AF063690; AAD50327.1; JOINED.  
 DR EMBL: AF063691; AAD50327.1; JOINED.  
 DR EMBL: AF063692; AAD50327.1; JOINED.  
 DR MGI: 1277201; COL13a1.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 9.  
 KW Collagen.  
 SQ SEQUENCE 751 AA; 73173 MW; FBE2443E1CBF51AD CRC64;

Query Match 64.1%; Score 59; DB 11; Length 751;  
 Best Local Similarity 73.3%; Pred. No. 1.1;  
 Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMGPNKDGKVG 15  
 DB 514 PAGPGPKGDKGPPG 528

RESULT 12  
 OY 09R41 PRELIMINARY; PRT; 1017 AA.  
 AC 09R41;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE SIMILAR TO ELASTIN MICROFIBRIL INTERFACE LOCATED PROTEIN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]



```

RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR. MAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
RC TISSUE.:
RA Strausberg R.:
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC005481; AAH05481.1; -.
DR InterPro: IPR001073; Clg.
DR InterPro: IPR000087; Collagen.
DR Pfam: PF00386; Clg; 1.
DR Pfam: PF01391; Collagen; 1.
DR SMART: SM0110; Clg; 1.
SQ SEQUENCE 1017 AA; 107584 MW; 2EF903204DB9C88F CRC64;

Query Match
Best Local Similarity 73.3%; Score 59; DB 11; Length 1017;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPMPGPKDGKVG 15
DB 834 PAGPMPGPKDGKVG 848
|||||

```

```

OC Paracentrotus.
OX NCBI_TaxID=7656;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90216744; PubMed=2324112;
RA D'Alessio M., Ramirez F., Suzuki H.R., Solursh M., Gambino R.;
RT "Cloning of a fibrillar collagen gene expressed in the mesenchymal
RT cells of the developing sea urchin embryo.";
RL J. Biol. Chem. 265:7050-7054(1990).
DR EMBL: J05422; AAA29440.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 13.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM0038; COLFI; 1.
KW Collagen.
FT NON_TER.
FT CHAIN
SQ SEQUENCE 1051 AA; 101961 MW; 81295E33E82727C7 CRC64;

Query Match
Best Local Similarity 62.0%; Score 57; DB 5; Length 1051;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 PAGPMPGPKDGKVG 15
DB 485 PAGPMPGPKDGKVG 499
|||||

```

```

RESULT 15
ID O90YJ0 PRELIMINARY; PRT: 1352 AA.
AC O90YJ0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROCOLLAGEN TYPE I ALPHA 2 CHAIN.
GN COL1A2.
OS Brachydonto rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Osteichthys;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Morvan-Dubois G., Le Guellac D., Garrone R., Zylberberg L.,
RA Bonnaud L.;
RT "Phylogenetic analysis on vertebrates fibrillar collagen enlightes
RT zebrafish alpha2(I) position and evidenced an evolutionary link between
RT collagen alpha chains and Hox clusters.";
RT Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Morvan-Dubois G., Haftek Z., Crozet C., Garrone R., Le Guellac D.;
RT "Structure and expression of the full length cDNA encoding zebrafish
RT alpha 2 type I collagen.";
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ318213; CAC51030.1; -.
KW Collagen.
SQ SEQUENCE 1352 AA; 127333 MW; 63D73C2CEDA44FE CRC64;

Query Match
Best Local Similarity 73.3%; Score 57; DB 13; Length 1352;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPGPKDGKVG 15
DB 1045 PAGPMPGPKDGKVG 1059
|||||

```

```

RESULT 16
O63123

```

```

ID 063123 PRELIMINARY; PRT; 1419 AA.
AC 063123; 063565;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE COLLAGEN ALPHA 1 TYPE II (T1 mRNA).
GN T1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BONE FRACTURE CALLUS;
RA Urahe K., Sarkar G., Bolander M.E.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1372-1419 FROM N.A.
RA Wurtz T., Brandsten C., Lundmark C., Christersson C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 370-422 FROM N.A.
RC STRAIN-DA: TISSUE-CARTILAGE;
RX MEDLINE=94321934; PubMed=8046350;
RA Michaelson E., Malmstrom V., Reis S., Engstrom A., Burkhart H.,
RA Holmdahl R.;
RT "T cell recognition of carbohydrates on type II collagen.";
RL J. Exp. Med. 180:745-749(1994).
DR EMBL; LA8440; AAA79780.1; -.
DR EMBL; AJ224879; CA12179.1; -.
DR EMBL; X79816; CA56213.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib-collagen_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 1.
DR Prodom: PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
SQ SEQUENCE 1419 AA; 134570 MW; B7C63B77819CE50B CRC64;

Query Match
Best Local Similarity 73.3%; Score 57; DB 11; Length 1419;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15
DB 848 PAGPMPNGKDGKVG 862

RESULT 17
062031 PRELIMINARY; PRT; 1442 AA.
AC 062031;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRO-ALPHA-1 TYPE II COLLAGEN.
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57/BLACK;
RX MEDLINE=91358489; PubMed=1885613;
RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT structure, and alternative splicing.";
RL J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C57/BLACK;
RA Vuorio E.;

```

```

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; M65161; AAA68099.1; -.
DR MGD; MGI:88452; COL2A1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib-collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00093; VWC; 1.
DR Prodom: PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Collagen.
SQ SEQUENCE 1442 AA; 137911 MW; D3A3274493CB821C CRC64;

Query Match
Best Local Similarity 73.3%; Score 57; DB 11; Length 1442;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15
DB 871 PAGPMPNGKDGKVG 885

RESULT 18
062033 PRELIMINARY; PRT; 1442 AA.
AC 062033;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRO-ALPHA-1 TYPE II COLLAGEN.
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C57/BLACK;
RX MEDLINE=91358489; PubMed=1885613;
RA Metzaranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT structure, and alternative splicing.";
RL J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C57/BLACK;
RA Vuorio E.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; M65161; AAA68102.1; -.
DR MGD; MGI:88452; COL2A1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib-collagen_C.
DR InterPro: IPR001007; VWFC.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 1.
DR Pfam: PF00093; VWC; 1.
DR Prodom: PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Collagen.
SQ SEQUENCE 1442 AA; 137829 MW; F0E77C11BCAFA93B CRC64;

Query Match
Best Local Similarity 73.3%; Score 57; DB 11; Length 1442;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPNGKDGKVG 15
DB 871 PAGPMPNGKDGKVG 885

```

```

RESULT 19
ID 062032 PRELIMINARY; PRT: 1459 AA.
AC 062032;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PRO-ALPHA-1 TYPE II COLLAGEN.
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6K;
RX MEDLINE=91358489; PubMed=1885613;
RA Metzaranta M., Tomán D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
  structure, and alternative splicing.";
RL J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57/BL6K;
RA Vuorio E.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; M65161; AAA68101.1; -.
DR MGD; MGI:88452; Col2a1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001007; VWFc.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; Vwc; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; Vwf; 1.
DR PROSITE; PS01208; VWFc; 1.
KW Collagen.
SQ SEQUENCE 1459 AA; 139071 MW; A09D24BF7357C827 CRC64;

Query Match 62.0%; Score 57; DB 11; Length 1459;
Best Local Similarity 73.3%; Pred. No. 4.4;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPGNKDGKYG 15
DB 888 PAGPMPGNKDGKYG 902

RESULT 20
ID 093208 PRELIMINARY; PRT: 303 AA.
AC 093208;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE C15A11.1 PROTEIN.
GN C15A11.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC Gardner A.E.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;

```

```

RT "Genome sequence of the nematode C. elegans: A platform for
  investigating biology.";
RT Science 282:2012-2018(1998).
DR EMBL; Z79694; CAB01959.1; -.
DR InterPro: IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 1.
SQ SEQUENCE 303 AA; 27889 MW; 9024926E86E07013 CRC64;

Query Match 60.9%; Score 56; DB 5; Length 303;
Best Local Similarity 66.7%; Pred. No. 1.2;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPMPGNKDGKYG 15
DB 210 PAGPMPGNKDGKYG 224

RESULT 21
ID 090E6 PRELIMINARY; PRT: 588 AA.
AC 090E6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE LATENT MEMBRANE PROTEIN 1 (FRAGMENT).
GN LMP1.
OS Cynomolgus Epstein-Barr Virus TSB-B6.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=104240;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90250747; PubMed=2160013;
RA Fujimoto K., Terato K., Miyamoto J., Ishiko H., Fujisaki M., Cho F.,
  Honjo S.;
RL "Establishment of a B-lymphoblastoid cell line infected with Epstein-
  Barr-related virus from a cynomolgus monkey (Macaca fascicularis).";
RL J. Med. Primatol. 19:21-30(1990).
RN [2]
RP SEQUENCE FROM N.A.
RA Faucher S., Wright K.E.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF181716; AAD56946.1; -.
DR InterPro: IPR002952; EggsHELL.
DR InterPro: IPR000188; GABA_receptor.
DR Pfam; PF02932; Neur_chan_memb; 1.
DR PRINTS; PR01228; EggsHELL.
FT NON_TER 588
SQ SEQUENCE 588 AA; 59838 MW; 62D85879901749BD CRC64;

Query Match 60.9%; Score 56; DB 12; Length 588;
Best Local Similarity 60.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 PAGPMPGNKDGKYG 15
DB 470 PAGPMPGNKDGKYG 484

RESULT 22
ID 090796 PRELIMINARY; PRT: 888 AA.
AC 090796;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ALPHA-1 TYPE XI COLLAGEN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OX NCBI_TaxID=9031;
RN [1]

```

RP SEQUENCE FROM N.A.  
 RC STRAIN-WHITE LEGHORN; TISSUE=LIMB;  
 RX MEDLINE=93054557; PubMed=1429607;  
 RA Nah H.-D., Barembaum M., Upholt W.B.;  
 RT "The chicken alpha 1 (X1) collagen gene is widely expressed in  
 embryonic tissues.";  
 RL J. Biol. Chem. 267:22581-22586(1992).  
 DR EMBL; M88593; AAA48707.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLF1.1.  
 DR Pfam: PF01391; Collagen\_10.  
 DR Prodom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLF1; 1.  
 KW Collagen.  
 FT NON\_TER  
 SQ SEQUENCE 888 AA; 86409 MW; B2ECD9A0E20F903E CRC64;

Query Match 60.9%; Score 56; DB 13; Length 888;  
 Best Local Similarity 66.7%; Pred. No. 3.7;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 DB 308 PPGPGGPGNGADGPGG 322

## RESULT 23

O44367 PRELIMINARY; PRT; 922 AA.  
 AC 044367;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PRECOLLAGEN D.  
 GN PRECOL-D.  
 OS Mytilus edulis (Blue mussel).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytilioida;  
 OC Mytilioida; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=6550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FOOT;  
 RX MEDLINE=98070444; PubMed=9405478;  
 RA Qin X.X., Coyne K.J., Waite J.H.;  
 RT "Tough tendons. Mussel byssus has collagen with silk-like domains.";  
 RL J. Biol. Chem. 272:32623-32627(1997).  
 DR EMBL; AF029249; AAB86638.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 7.  
 KW Collagen.  
 SQ SEQUENCE 922 AA; 80306 MW; 599D155E47A2C24A CRC64;

Query Match 60.9%; Score 56; DB 5; Length 922;  
 Best Local Similarity 66.7%; Pred. No. 3.8;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 DB 444 PPGPGGPGGDEGPGV 458

## RESULT 24

O26154 PRELIMINARY; PRT; 1076 AA.  
 AC 026154;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE V-SERA 3.  
 GN V-SERA 3.  
 OS Plasmodium vivax.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI\_TaxID=5855;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SALVADOR I;  
 RX MEDLINE=96408670; PubMed=8813677;  
 RA Kieffer M.C., Crawford K.A., Boley L.J., Landsberg K.E., Gibson H.L.,  
 RT Kaslow D.C., Barr P.J.;  
 RT "Identification and cloning of a locus of serine repeat antigen  
 (sera)-related genes from Plasmodium vivax.";  
 RL Mol. Biochem. Parasitol. 78:55-65(1996).  
 DR EMBL; U51723; ABA41487.1; -;  
 DR InterPro: IPR001064; Crystallin.  
 DR InterPro: IPR000668; Peptidase\_C1.  
 DR Pfam: PF00112; Peptidase\_C1; 1.  
 DR PRINTS: PR00705; PAPA1N.  
 DR PROSITE: PS00225; CRYSTALLIN\_BETAGAMMA; UNKNOWN\_1.  
 SQ SEQUENCE 1076 AA; 11597 MW; 0B8C2509C3CCT1AB CRC64;

Query Match 60.9%; Score 56; DB 5; Length 1076;  
 Best Local Similarity 66.7%; Pred. No. 4.5;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 DB 843 PDGAPGSGPDGNGV 857

## RESULT 25

O9W7R9 PRELIMINARY; PRT; 1418 AA.  
 AC O9W7R9;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALPHA TYPE II COLLAGEN  
 OS Cynops pyrrhogaster (Japanese common newt).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Caudata; Salamandroidae; Salamandridae; Cynops.  
 OX NCBI\_TaxID=8330;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99407244; PubMed=10474166;  
 RA Asahina K., Obara M., Yoshizato K.;  
 RT "Expression of genes of type I and type II collagen in the formation  
 and development of the blastema of regenerating newt limb.";  
 RL Dev. Dyn. 216:59-71(1999).  
 DR EMBL; AB022046; BAA82043.1; -;  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000885; Fib\_collagen\_C.  
 DR Pfam: PF01410; COLF1; 1.  
 DR Pfam: PF01391; Collagen; 18.  
 DR Prodom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLF1; 1.  
 KW Collagen.  
 SQ SEQUENCE 1418 AA; 135067 MW; C19A6E601A2A717E CRC64;

Query Match 60.9%; Score 56; DB 13; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 6;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPMPNGKDGKVG 15  
 DB 1096 PPGPVGPGSGKDGSG 1110

## RESULT 26

O9OW37 PRELIMINARY; PRT; 1420 AA.  
 AC O9OW37;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALPHA 1 TYPE IIA COLLAGEN PRECURSOR.

```
GN COL2A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
ON NCBI_TaxID=9031;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERMAL;
RA Calixia X., Yongzhi X., Sidi G., Yiyang S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF046949; AAK98621.1; -.
KW SIGNAL; Collagen.
FT SIGNAL.
SQ SEQUENCE 1420 AA: 13499 MW: 8899AAB17F214FF5 CRC64;

Query Match 60.9%; Score 56; DB 13; Length 1420;
Best Local Similarity 66.7%; Pred. No. 6;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PACPGPGRKDGKGVG 15
Db 1098 PGPVGPGRKDGSGNG 1112

RESULT 27
O91717 PRELIMINARY; PRT; 1486 AA.
AC O91717;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ALPHA-1 TYPE II COLLAGEN.
GN COL2A1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92011898; PubMed=1918153;
RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.;
RT "Expression of two nonallelic type II procollagen genes during Xenopus
RT laevis embryogenesis is characterized by stage-specific production of
RT alternatively spliced transcripts."
RL J. Cell Biol. 115:565-575(1991).
DR EMBL: M63595; AAA49678.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; WFC.
DR Pfam: PF01410; COLFI.1.
DR Pfam: PF01391; Collagen_18.
DR Pfam: PF00093; vwc; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00214; VWC; 1.
DR SMART: SM00038; COLFI.1.
DR PROSITE: PS01208; VWF; 1.
KW COLLAGEN.
SQ SEQUENCE 1486 AA: 142263 MW: 4A4A95772341042F CRC64;

Query Match 60.9%; Score 56; DB 13; Length 1486;
Best Local Similarity 66.7%; Pred. No. 6.3;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
```

```
AC O91718;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE ALPHA-1 TYPE II COLLAGEN.
ON NCBI_TaxID=8355;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=92011898; PubMed=1918153;
RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.;
RT "Expression of two nonallelic type II procollagen genes during Xenopus
RT laevis embryogenesis is characterized by stage-specific production of
RT alternatively spliced transcripts."
RL J. Cell Biol. 115:565-575(1991).
DR EMBL: M63596; AAA49679.1; -.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR000885; Fib_collagen_C.
DR InterPro: IPR001007; WFC.
DR Pfam: PF01410; COLFI.1.
DR Pfam: PF01391; Collagen_18.
DR Pfam: PF00093; vwc; 1.
DR ProDom: PD002078; Fib_collagen_C; 1.
DR SMART: SM00038; COLFI.1.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS01208; VWF; 1.
KW COLLAGEN.
SQ SEQUENCE 1491 AA: 142495 MW: 43026FF08FB0314 CRC64;

Query Match 60.9%; Score 56; DB 13; Length 1491;
Best Local Similarity 66.7%; Pred. No. 6.3;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PACPGPGRKDGKGVG 15
Db 1170 PGPVGPGRKDGSGNG 1184

RESULT 29
O90UT4 PRELIMINARY; PRT; 1767 AA.
ID O90UT4;
AC O90UT4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE COLLAGEN TYPE XI ALPHA-1.
GN COL11A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20455728; PubMed=10486316;
RA Annunen S., Korhko J., Czarny M., Warman M.L., Brunner H.G.,
RA Kaerlainen H., Mulliken J.B., Tranebjaerg L., Brooks D.G., Cox G.F.,
RA Cruysberg J.R., Curtis M.A., Davenport S.L.H., Friedrich C.A.,
RA Kaitila I., Krawczynski M.R., Latos-Bielska A., Mukai S.,
RA Olsen B.R., Shimo N., Somer M., Viikula M., Zlotogora J.,
RA Prockop D.J., Ala-Kokko L.;
RT "Splicing Mutations of 54 bp Exons in the COL11A1 Gene Cause Marshall
RT Syndrome but Other Mutations Cause Overlapping Marshall/Stickler
RT Phenotypes."
RL Am. J. Hum. Genet. 65:974-983(1999).
DR EMBL: AF101112; AAF04724.1; -.
DR EMBL: AF101079; AAF04724.1; JOINED.
DR EMBL: AF101080; AAF04724.1; JOINED.
DR EMBL: AF101081; AAF04724.1; JOINED.
DR EMBL: AF101082; AAF04724.1; JOINED.
DR EMBL: AF101083; AAF04724.1; JOINED.
```

